CAPTURING 'AHA!' MOMENTS OF PUZZLE PROBLEMS USING PUPILLARY RESPONSES AND BLINKS

by

Tsunhin John Wong

Bachelor of Cognitive Science, The University of Hong Kong, 1999

Master of Philosophy, The University of Hong Kong, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science in Cognitive Psychology

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This thesis was presented

by

Tsunhin John Wong

It was defended on

July 16th, 2009

and approved by

Timothy J. Nokes, PhD, Assistant Professor, Department of Psychology

Christian D. Schunn, PhD, Associate Professor, Department of Psychology

Greg J. Siegle, PhD, Associate Professor, Department of Psychiatry

Thesis Director: Christian D. Schunn, PhD, Associate Professor, Department of Psychology

CAPTURING 'AHA!' MOMENTS OF PUZZLE PROBLEMS USING PUPILLARY RESPONSES AND BLINKS

Tsunhin John Wong, M.S.

University of Pittsburgh, 2009

Inspired by the design of Metcalfe and Wiebe (1987), this study is a quest for more fine-grained measures of the differences between insight problem solving and non-insight problem solving at the moment of an 'Aha!'. We avoided asking for metacognitive reports that have been shown to disturb insight problem solving (Schooler, Olsson, & Brooks, 1993), and made use of routine (non-insight) problems of a very similar task nature to compare with the performance, eye-related physiological measures, and subjective ratings of participants on their solving of puzzle (insight) problems. We found that puzzle problems and routine problems led to different pupillary dilations with perception of the tasks and mood as possible mediators of the effect. Profiles of pupillary responses and blinks are found to be indicators that depict characteristics of the task nature when a person is unable to predict his/her own problem solving performance on an insight problem. We also found that discontinuity and restructuring are possibly more important in characterizing an insight problem solving event. Third, instead of the traditional sharp and abrupt but indistinct characterization of the insight moment, our data suggested that an insight is possibly an emerging moment of a sense of clarity about the solution which takes time before the time of an 'Aha!' solution declaration, which converges with some of the claims made according to records of historical scientific discoveries (Gruber, 1995).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I want to thank my advisor Dr. Christian Schunn and my co-advisor Dr. Greg Siegle for their patience and effort to help me complete this project. I thank Tim Nokes for his comments that inspired me to restructure the framework of this study to this current state. Finally, I thank my parents for their patience and understanding to allow me to purchase a career in the academia.

1.0 INTRODUCTION

It is believed that one day when Archimedes was bathing, he suddenly realized that the volume of an irregular object could be calculated from the volume of water displaced when the object was submerged in water. He cried ‘Eureka!’ repeatedly, leapt out of his bathtub, and ran in excitement through the streets of Syracuse forgetting to dress. This is a classic case of gaining insight during problem solving – and an example of insight problem in which the problem solver has to change his/her perspective, coming to see the problem in a novel way in order to reach a solution.

In psychology, there is a similar characterization of insight-problem solving by Metcalfe and Wiebe (1987) of insight as a sudden unforeseen flash of illumination. Their study took a closer look at our metacognitive ability towards insight problems versus noninsight problems. They hypothesized that not being able to be monitored is one key characteristic of insight problems. In their study, they contrasted insight problems with noninsight problems that need incremental problem solving steps. Participants were shown insight and noninsight problems one-by-one, and they had to indicate their subjective proximity to the solution by every 15 seconds. Metcalfe and Wiebe also had participants provide subjective performance predictions for the insight and math problems. With this method, they explored metacognitive differences between insight problems versus noninsight problems. Their basic findings were that participants were not able predict their own performance on insight problems, and probability and distribution based normative data provide better predictions on individual performance than their own predictions. The “warmth” ratings, which are ratings to indicate the subjective proximity to a solution, were also quite different between the insight problems and noninsight problems. While the ratings show an incremental progress for the noninsight problems, the ratings for insight problems showed little premonition of the impending insight (see figure 1).

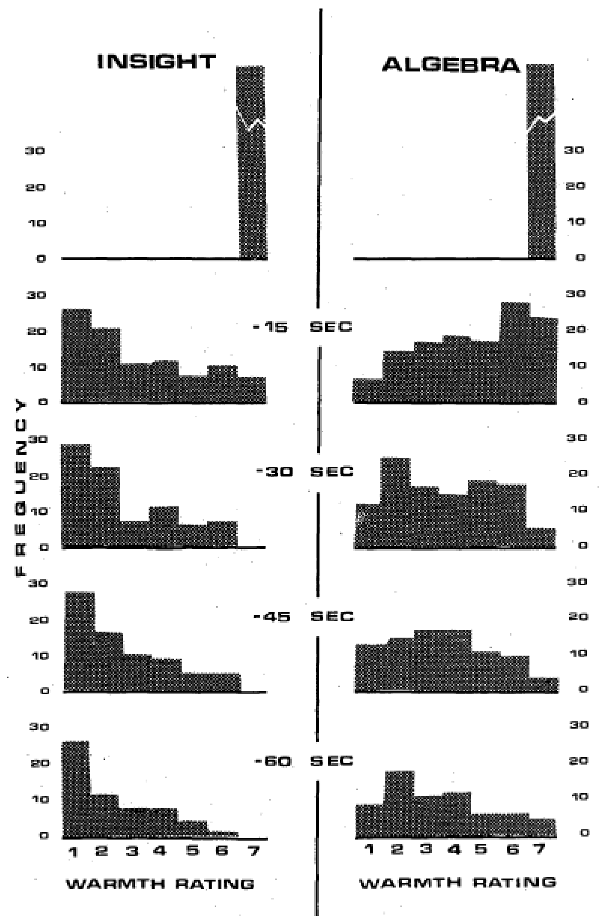


Figure 1. Warmth ratings of insight problems and noninsight problems at 15 second intervals, ranging from 60 seconds prior to solution to the solution moment. (Reprinted from Metcalfe & Wiebe 1987)

Using this approach of having participants provide “warmth ratings” to the proximity to solution, they are among the first researchers to demonstrate that insight problem solving is different from non-insight problem solving in that problem solvers are unable to foretell the coming of an insight solution. Their finding also signified that insight problems are less open to be analyzed by our conscious metacognitive skills that, by contrast, can be easily applied to noninsight problems. However, as data was collected only in 15-seconds intervals, the design may have missed events occurring in that 15-seconds right before the burst of insight.

To further complete the story of insight versus noninsight problem, the current study looks in more detail at the few seconds before and after a person has just reached the solution for a problem. To serve the purpose of monitoring problem solvers, we have to select means to monitor problem solvers online and with a fine-grained temporal resolution. Also we seek methodologies other than self-report as previous studies have shown that there are nonreportable aspects of insight problem solving (Metcalfe & Wiebe, 1987; Schooler, Ohlsson, & Brooks, 1993). We have selected eye-tracking to meet these requirements. Apart from the online and fine-grain data sampling nature of eye-tracking, the next section explains why eye-tracking is utilized here.

1.1 COGNITIVE LOAD AND AROUSAL LEVEL DURING PROBLEM SOLVING ACTIVITIES

Insight and non-insight problem solving are thought to differ in terms of cognitive load and arousal, and both of those factors can be measured using eye-tracking.paragraph.

1.1.1 Cognitive load

Cognitive load in the problem solving literature refer to working memory load and process execution load. In 1974, a popular model of working memory was proposed by Baddeley and Hitch (1974) suggesting that our short-term memory system is comprised of multiple components—a phonological loop, a visual-spatial sketchpad, an episodic buffer, and a central executive—that are all capacity limited. There are a number of factors that govern working memory load and process execution load. For example, goal and subgoal structure of a problem affects goal retention and retrieval (Altmann & Trafton, 2002; Anderson & Douglass, 2001), the amount of content a problem solver needs to internalize (Anderson, Reder, & Lebiere, 1996), and the number of intermediate steps that a person needs to hold for the problem (Ericsson & Kintsch, 1995).

Variations of cognitive load and memory capacity occur in many theories about human problem solving. Models using cognitive architectures (e.g. ACT-R) can make very accurate predictions on many tasks based on these capacity limitations. For example, when a task has a well-defined goal, and the problem solver knows the proper solution path, then time and accuracy of performance as well as difficulty of the task can be predicted based on the complexity of solution steps and subgoal structure, the frequency and amount of working memory he/she needs to access, and the content as well as the sophistication that a piece of knowledge information that he/she needs to retrieve. However, these predictions are mainly about noninsight problem solving, leaving the realm of insight problems less explored. Whether we can apply all the theoretical components and assumptions of working memory capacity and process limitations to insight problem is less straightforward.

There are three major theoretical perspectives on insight problem solving, and their theories each have implications for cognitive load during the course of problem solving. They will each be described in the next sections.

1.1.1.1 Problem solving by analogies

Gick and Holyoak (1980, 1983), and Gentner (1983) characterize analogical (insight) problem solving as realizing the relevance of previous examples to present task, and mapping of problem elements between previous example and current task at hand. In the early 80s, Gick and Holyoak (1980, 1983) made use of Duncker's problem (Duncker, 1945) to derive a theory of how humans make use of analogies to solve novel problems. Their theory involves schema formation for induction and transfer to novel problems later on. At approximately the same time, Gentner (1983) proposed a theory to explain how an analogy is utilized by having its elements mapped onto the target scenario or problem. These two streams of theories on analogy became more compatible following some collaborative work (Gentner, Holyoak, & Kokinov, 1998).

Their studies made use of a number of insight problems in which problem solvers encounter sudden transitions to the solution state. Also, they have explained the mechanisms underlying scientific discovery through the use of analogical mechanisms in metaphors for discoveries (Holyoak & Thagard, 1997; Namy & Gentner, 2002). Christensen and Schunn (Christensen & Schunn, 2005, 2007) have shown that analogies play a role in insight problems, and are useful in real world creativity settings. In all of their theories, they have a focus on how relational patterns work so that pre-existing knowledge can contribute to insight problem solving in novel scenarios.

Within this theoretical perspective, we predict cognitive load during insight problem solving based on the MAC/FAC model of Gentner et. al. (Gentner, Forbus, & Law, 1995).

MAC/FAC model is a two stage similarity-based case retrieval model: 1) a computationally economical MAC phase that filters out unlikely candidates, and 2) a more computationally expensive FAC stage that performs more detailed mapping and evaluative matching processes for candidates that have passed through the filter. The FAC stage is based on structural mapping engine (Falkenhainer, Forbus, & Gentner, 1989), and is most suitable for base and target problem pairs that have a clear structure for mapping. The model posits that the filtering phase (MAC phase) is cognitively inexpensive, and therefore more cognitive load should occur during the FAC phase in which selected examples are structurally mapped and compared among themselves. The model also predicts that cognitive load should be higher when there are more matching examples that can pass into the expensive selector in the FAC stage. A possibly high demanding phase in the case retrieval process would be at this late matching phase where previous cases are mapped and compared to the target, and when a good match is just about to be pulled out for an insight to occur (See figure 2). However, their model can only apply to problems where structural contents are present.

1.1.1.2 Search Heuristic in the Problem Space

In 1990, by extending the original problem space theory (Newell & Simon, 1972), Kaplan and Simon (1990) approached the topic of insight problem solving. They posited that the ability to notice invariants in a problem space is crucial for getting out of a mental rut. Lovett and Schunn (1999) also proposed that search for an alternative representation of a problem happens when current problem solving strategies are not successful. Later on, MacGregor et. al. (MacGregor, Ormerod, & Chronicle, 2001) develop a theory that further elaborated how search heuristics impact problem solving performance when tackling insight problems. The basic idea is that insight problems trick the problem solvers into using heuristics inadequate for the problem at hand. Heuristics, as they described them, are content-independent strategies that allow one to make progress on a broad range of problems. When problem solvers use heuristics that are inadequate for the problem at hand, they encounter impasses. Insight happens when problem solvers figure out the appropriate heuristic to approach that particular insight problem. Rather than questioning how knowledge of a solution is located or formed, this school of thought focuses on a meta-cognitive level of control during problem solving.

In their framework, working memory is required to execute metacognitive strategies utilized by problem solvers to monitor progress. Judgment processes about whether current heuristics are working or not, as well as shifting to newer heuristics also impose a processing load on problem solvers. Also, Simon and Chase's (1973) idea of chunking during the search for a solution in the solution space has additional implications for the working memory load – the bigger the chunks a problem solver can use to encode the situation, the less working memory a problem solver is going to use during problem solving. Because their theories are developed from multiple-steps problems, they could examine whether problem solvers were using highly demanding strategy of mental lookahead or a criterion-based probability judgment on the current solution step. A switch in heuristic is planned or executed at the time when a criterion judgment for a current move suggests that the present strategy is getting the problem solver nowhere and that the solution search is at an impasse. Various kinds of processes are occurring at the same time during the switch in heuristic – monitoring of success rates, probability judgments for whether the current move is failing to meet the success criterion, and a constant search for alternative rational moves during the problem space. Therefore it is probably the time with the highest cognitive load. In their nine-dot problem experiments (MacGregor et al., 2001), predictions for probability of success are all based on possible lookahead within systematically modified problems (See figure 3). However, the authors admitted that their progress-monitoring heuristic for insight problems is likely to be inappropriate for insight problems that have only a single step, for example the matchstick algebra problems used by Knoblich et al. (Knoblich, Ohlsson, Haider, & Rhenius, 1999). While, their models make predictions of success mainly based of possible lookaheads in multiple-step problems, the lookahead also predicts difficulty and cognitive load of a problem solver – the more steps involved in a lookahead, the higher the cognitive load a solver will experience. Apart from that, their theory implies a constant load in terms of seeking a possible alternative solution step at the particular stage of problem (i.e., a locally rational alternative operator).

1.1.1.3 Restructuring of Problem Representation

The theoretical approach that conceives of insight problem solving as restructuring of representations can be regarded as a recent revival of notions from Gestalt psychology. The Gestalt psychologists were among the earliest investigators of the insight phenomenon (Köhler, 1921; Wertheimer, 1925). They hypothesized that the restructuring that occurs during sudden transitions in insight problem solving is similar to that which happens during the perception of ambiguous figures like the Necker cube. Ohlsson and Knoblich (Knoblich et al., 1999; Ohlsson, 1984, 1992) suggested that insights are the alterations of problem representations to make available additional useful problem solving moves. They have further proposed cognitive processes by which problem representations are restructured. Re-encoding (Ohlsson, 1992) is essential for the problem solving scenario where the problem elements are interpreted in a manner that prevents the solution. They name the second proposed process of representational changes as “chunk decomposition” (Knoblich et al., 1999), which describes the need to divide or transform the encoded problem elements, which are inadequate for a solution to be performed, into smaller semantic units in order to allow for the appropriate solution steps to be performed. The third proposed process, named “constraint relaxation” (Knoblich et al., 1999; Ohlsson, 1992), involves a change in the representation of the goal of problem solving which is usually overly narrow and biased by prior knowledge. This theoretical approach of problem restructuring stresses that problem solvers achieve insights through changing perspectives and discovering the immediate operators by which the solution is achieved.

According to their characterization of insight problem solving, these re-encoding processes known as chunk decomposition and constraint relaxation certainly impose a cognitive load. Recognition of which chunk is to be decomposed and which constraints are to be relaxed are processes at the moment of insight, while there are also components of access and retrieval of memory for the matching patterns of decomposed chunks and relaxed constraints. Their studies made predictions for success rate, and solution time based on the difficulty of constraints to be relaxed and chunks to be decomposed in order to reach their single-step solutions. Their framework does not focus on working memory load and processing load, except that they predict a state of “kaleidoscopic confusion or stimulus overload” during the time when a problem solver enters an impasse. Therefore, there is the implication of having a working memory overload during the impasse moment (See figure 4). Note that these predictions (and all other prediction graphs in this thesis) are meant to be qualitative/heuristic in nature rather than quantitative predictions. Also, as their theory describes mainly the key single step process through which an impasse is resolved and hence an insight is acquired, there is no prediction for the time of highest cognitive load, except that initiation of impasses and then resolution of impasses (acquisition of insights) are the most prominent processes in insight problem solving.

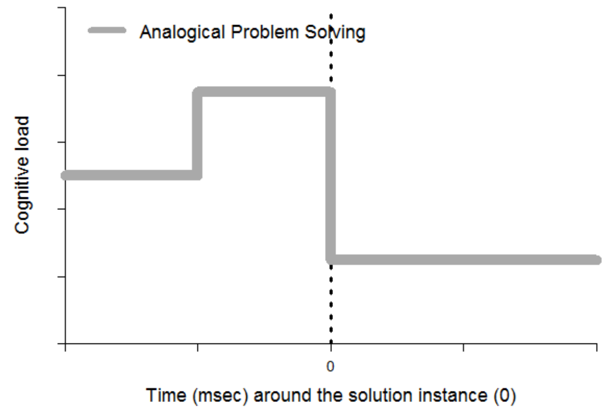


Figure 2. Temporal progression of cognitive load according to MAC/FAC.

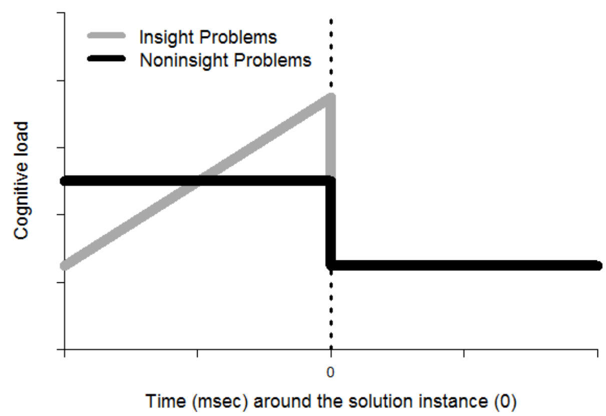


Figure 3. Temporal progression of cognitive load according to search heuristic theory.

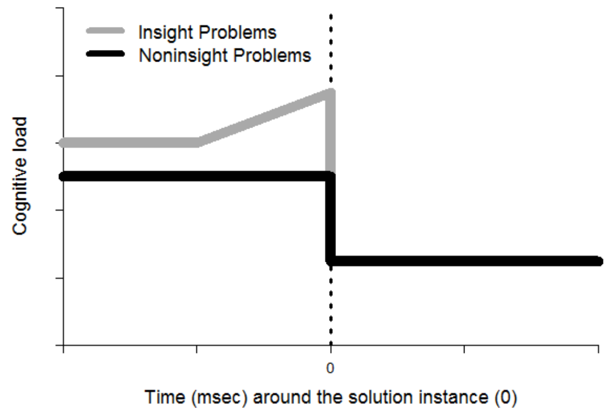


Figure 4. Temporal progression of cognitive load according to restructuring theory.

1.1.2 Affective components

When problem solving tasks are imposing a working memory load and cognitive process load on a solver, the stress from these loads affect the arousal level of the solver. Arousal can be defined as the level of physiological reactivity of a person (Broadbent, 1971; Eysenck, 1982; Kahneman, 1973). Sleep is at one end of the spectrum and an excited or panicked mode is at the other end of it. The fluctuation of arousal level is proposed to be modulated by a dual-level control mechanism (Broadbent, 1971; Eysenck, 1982). The low-level passive physiological arousal system is actively modulated by a higher-level cognitive arousal system. Fluctuation in arousal level of the low-level passive system, due to sleep deprivation for example, will be compensated by the high-level system in an attempt to maintain a person in an optimal level of arousal for performing problem solving.

It takes time and effort to execute a cognitive process. For example, pattern matching for possible actions, conflict resolution among possible actions, and those computations that a person needs to perform during search, retrieval and application of relevant knowledge in the process of problem solving as suggested by some of the cognitive architecture theories (Anderson, 1993). These processes take time and effort to execute, and a person has limitations on how fast these processes can be carried out. The more demanding a problem solving task is, the higher the arousal level that will be induced. In addition to arousal level, emotions and problem solving are also related through the notions of motivation and confidence (e.g. self-efficacy), although those elements are beyond the scope of the current study.

While arousal is one important part of the emotional reactions of a problem solver, researchers have also noted other aspects of emotion during insight problem solving. Gick and Lockhart (1995) argue that affective components have an important role in characterizing insight problems. They described the phenomena of “Aha!” during insight problem solving as having two major components. The first feature is the surprise and the unexpectedness as the person solves the problem in a very different way. The second feature is the suddenness because the solution is often easily applicable in an obvious way. Other researchers (Dominowski & Dallob, 1995) have also described the suddenness of the emergence of solution as an important characteristic of insight problem solving. Köhler (1927) also stressed the sudden changes of behaviour of apes in their studies of primate problem solving. After careful examinations of historical and biographical accounts of some the most celebrated cases of insights (e.g., Archimedes’ Bath, Kekule’s Dreams, Einstein’s 10-years Phases, Darwin’s Malthusian Moment, Piaget’s Joy), Gruber (1995) concluded that certain stages of development of these scientific theories are subjectively much bigger and more sudden, and they brought joy and relief to these scientists amidst the despair and frustration of impasses. Surprise and the feeling of ease and joy due to discovery of the solution are argued to be the most prominent emotions related to insight problem solving.

Ekman (1992) argues that basic emotions have nine characteristics which distinguish them from other more complex emotions and that these characteristics are important because they ensure that these emotions have a biological contribution to the organism. For example, a basic emotion is proposed to have distinctive universal facial expression, some emotion-specific physiology, common elements and contexts in which that emotion is found to occur. ‘Surprise’ is among those emotions that are considered incomplete in their characteristics to be regarded as a basic emotion in this system. Also, Ekman found that many positive emotions do not have a distinctive signal, resulting in a suggested grouping together of ‘amusement’, ‘relief’, ‘sensory pleasure’, ‘pride’, ‘excitement’, ‘satisfaction’, and ‘contentment’.

In light of these findings in emotion research, we are going to use a gloss of ‘arousal level’ to cover the sudden surprise of ‘Aha!’ that a problem solver encounters during insight problem solving. Broader terms of ‘enjoyment’ and ‘frustration’ are used in our overall evaluation of the positive and negative valence aspect of emotional experience during insight problem solving. Figure 5 is a visualized temporal progression of the affective components of frustration and enjoyment as suggested by the depictions of insight (aka ‘Aha!’) in the research of insight problem solving. That is, frustration will build and then release at the moment of insight, being replaced by a sudden spike in positive emotions, which then gradually dissipate. From a total arousal perspective, there is a gradual build, a spike at the moment of insight, and then a gradual decay.

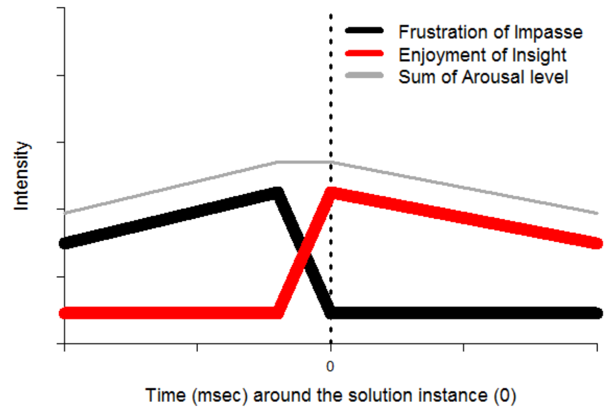


Figure 5. Temporal progression of affective components of 'frustration', 'enjoyment', and total arousal during problem solving, as characterized by theories and historical case studies about the 'Aha!' phenomena.

1.2 RELATIONS BETWEEN COGNITIVE LOAD AND AROUSAL LEVEL TO EYE-TRACKING DATA: PUPILLOMETRY AND BLINKS

Besides being able to provide finer-grain data in a synchronized (online) manner, there are potentially other advantages to utilizing eye-tracking to study insight problem solving. Prior problem solving studies that made use of eye tracking techniques (Chronicle, MacGregor, & Ormerod, 2004; Grant & Spivey, 2003; Knoblich, Ohlsson, & Raney, 2001) tended to focus on eye fixations and saccades, typically focusing on frequency of visits to defined regions of interest (ROI) within the stimuli. While fixation durations and saccades do provide important information about cognitive processing (Ballard, Hayhoe, Pook, & Rao, 1997), there is additional information in the pupillometry profiles and blink profiles that relate to the process of solving insight problems. Therefore, rather than examining fixation durations and locations, we are going to make use of pupillometry and blinks data to measure aspects of the insight phenomena during problem solving.

1.2.1 Pupillometry

While there has been a long history of research on the pupillary light reflex, there is also a considerable wealth of studies on the psychological side of pupil activities in terms of thinking and emotion. With regard to mental activity and pupil size, Janisse (1977) pointed out that interest in the topic dates back to the late 1800s. He credited Kahneman for having documented the phenomena in an organized fashion and in terms of cognitive expression, such as “loading”, “unloading”, “processing”, “effort”, and “rehearsal”. Using a digit-span task, Kahneman and Beatty (1966) suggested that the pupil provide an effective index of processing load, showing that pupil size varied as a function of both within-task and between-task changes in difficulty. In a review of various studies of pupillary responses and thinking, Janisse (1976) argued that pupil size served as a measure of mental effort. A more recent study (Beatty, 1982) also suggested that the pupil is larger when a person is performing more difficult tasks, with higher attentional and memory demands.

In terms of arousal, the earliest relevant ideas about the pupil involved the relationship of the pupil in response to changes in the autonomic nervous system. Darwin (1872) pointed out that pupillary dilation can be related to fear and other “emotions” in animals. Janisse (1977) reported a series of studies showing that pupil size, rather than reflecting emotional valence, may mainly convey the difference in intensity or strength between attitudes or emotional states. However, for task related arousal level, Kahneman (1973) argued that a measure of arousal is indeed a measure of the "demands" in components of a task, and therefore a reflection of "mental effort". And he stated that dilation of the pupil is a very reliable and consistent physiological index of effort (p. 18-24).

There are many other factors that may induce changes in pupil size. For example, Janisse (1976) summarized a number of different pupillometry literatures and characterized the pupil as "a response to a stressor", "related to certain scales of anxiety and related constructs", "a measure of sexual arousal", "a measure of anticipation before an expected event", and "a measure of mental effort" (see Janisse, 1976 for a more comprehensive review). There is even a phenomenon of pupillary unrest called the pupillary hippus, which involves the constant irregular dilation and constriction of pupil in any diffuse light, with a more exaggerated effect when the subject is fatigued. However, Kahneman (1971) pointed out that while the pupil of a resting subject is quite variable due to a number of factors, the pupil of an attentive subject engaged in a task is relatively very stable.

Pupillary responses do have a short lag after the time of initial neuronal reactivity. One study found that waveforms for a cued-reaction-time task peaked at about 0.5s, but note that their task took nearly 3s to resolve (Siegle, Granholm, Ingram, & Matt, 2001). Onset of dilations due to light can be as early as 300 ms after simulation with initial peaks at about 500 ms, usually followed by more extensive peaks at 1200 ms or later for cognitive and simple motor tasks. Similar lag profiles were found in pain-evoked responses, but these responses include both the fast neural as well as longer muscle/inertia based aspects; most of the direct information about lag profiles is from animal recordings, where the nerves can be isolated, and evoked potentials could be recorded from the afferent nerves or ganglia due to peripheral stimulation (Loewenfeld & Lowenstein, 1999).

1.2.2 Blinks

In addition to studies involving cognition and pupillary response, there are studies suggesting a relationship between eyeblinks and various cognitive activities. Spontaneous eyeblinks are thought to be cortically controlled, and reflect cognitive states relating to the allocation of attentional resources, transition points in information processing flow and processing mode (Stern, Walrath, & Goldstein, 1984). Fogarty and Stern (1989) argued that blink latencies or blink inhibitions are related to the importance of information acquired during a particular gaze, modulating the period of information acquisition. On the other hand, Ohira et. al. (Ohira, Winton, & Oyama, 1998) suggested that blink rate increases when there is a concurrent task that requires cognitive processing, for example judging or rating. They also suggested that valence of stimuli will affect eyeblink latency in different ways – longer latencies for negative stimuli than for positive stimuli.

Recent progress in the area (Ichikawa & Ohira, 2004; Siegle, Ichikawa, & Steinhauer, 2008) points out that the distribution of eyeblinks is related to the end of cognitive processing. Siegle, Ichikawa, and Steinhauer (Siegle et al., 2008) further suggest that eyeblinks occur in the early stage of sensory processing and also at the time when sustained information processing is finished. Through analyses of the time course of blink frequency alongside with pupil dilation in digit sorting tasks and Stroop tasks, their study also points out that eyeblink and pupil data are complementary to one another, with each describing different aspects and stages of information processing. Upon sensing an external stimulus, they found that there is an immediate increase in blink activities, and then a decrease, leading to a time of comparable levels of activities between blinks and pupillary dilations. Pupillary dilations continued to increase to a peak where their level of activities exceeds that of the blinks. And after processing of a task, pupils were observed to constrict, and the activities of blinks increased to exceed that of the pupil once again. Overall, eyeblinks appear in the early phase of cognitive processing and in the late or finishing stage of prolonged processing; an inhibition of eyeblinks is expected in the middle of cognitively demanding tasks.

In sum, pupillary responses and eyeblinks provide information on the fluctuations in cognitive states of problem solvers, in particular about their cognitive processing and cognitive load during the task. As the task becomes more cognitive demanding (e.g. taking up more working memory, or involving more complication computations), pupillary dilations are expected as a result of higher mental effort or cognitive load, as well as because of a heighten arousal level following. Eyeblinks should be less frequent during the information acquisition stage and the more intense cognitive processes, and should be more frequent before and after these relatively busier stages of a problem solving task. As a more unique feature of insight problem solving, pupillary dilations should follow the joy of relief and the surprise and suddenness feeling accompanied by the exhilaration of "Aha!" moment during resolution of impasse (and acquisition of insight), and more eyeblinks should follow when an impasse is resolved.

1.3 GENERAL DESIGN

In search of more fine-grained measures of the differences between insight problem solving and noninsight problem solving, we adopted a design similar to that of Metcalfe and Wiebe (1987). We presented insight problems and noninsight problems in randomized orders to our participants. However, as our focus is not on metacognition during problem solving, but rather on developing finer-grain measures of cognitive states in insight problem solving and noninsight problem solving, we made use of eye-tracking to monitor our participants during the task. Also, instead of having them give "warmth ratings" and scores on "feeling of knowing" in given time intervals, we asked them to give ratings about their moods during solving the problem and the insight-like-ness of their experience of solving the problem at the completion of each problem to provide some convergent validity to our eye-tracking measures.

Our study focuses on a few key predictions regarding the differences between puzzle problems and routine problems behaviors. First, contrary to the unreportable metacognitive aspect of insight problem solving (Metcalfe & Wiebe, 1987), we are expecting that there will be a characteristic peak of higher cognitive load and arousal level around the time of impasse resolution, i.e. burst of insight. Such a peak of arousal level and cognitive load at the time of solution discovery for a puzzle problem will be followed by a fall to a normal level of load and arousal. These fluctuations of cognitive load and arousal level will be denoted by pupillary dilations, and the terminations of such an intensive period of activity will be followed by a heightened frequency of eyeblinks. For routine problems, if there is any sizeable fluctuation in cognitive load and arousal level during problem solving, it should be less intense than those demonstrated during insight problem solving activities. Fluctuations in pupillary dilations and eyeblinks frequency with a task will therefore be less obvious than during the process of insight problem solving. Second, while Metcalfe and Wiebe (1987) posited from their results that metacognition cannot serve to predict insight outcomes, we are interested to know if there is any relationship between physiological measures (pupillary responses and eyeblinks) and subjective ratings of affect states during problem solving.

As in the study of Metcalfe and Wiebe (1987), we made use of routine problems that have an incremental nature, and that require a step-by-step approach to reach the solution. However, we replaced both the puzzle problems and route tasks with our own version of matchstick problems that we have adopted and developed from the examples utilized in the study of Knoblich et al. (1999) on insight problem solving. There are a few reasons to adopt the matchstick problems. First, these are single-step insight problems, so that the moment of insight can be robustly identified as the single-step of impasse resolution. Second, the puzzle problems of matchstick arithmetic can be systematically varied to different difficulty levels. By varying the level of difficulty, we are able to control the likelihood that participants reach impasses and experience an insight moment. Third, we can create comparable incremental routine problems by modifying matchstick puzzles. This approach avoids making comparisons between "apples and oranges", as many studies on insight problem solving have compared puzzle problems to considerably different forms of problems (e.g. involving problems with different conceptual elements and different visual complexity). This issue of comparable components is especially important for a study of the fine-grained time course of insight and noninsight problem solving. Also, although we are quite sure the match stick puzzle problems utilized by Knoblich et al (Knoblich et al., 1999; Knoblich et al., 2001) are going to resemble some key characteristics of insight problems, we are going to call our tested insight problems "puzzle problems", and our tested noninsight problem "routine problems" because we do not want to make an assumption that our participants must experience a moment of insight during solving of these puzzle problems. Instead, we introduced some subjective ratings for our participants to judge if the problem solving experience on these puzzle problems have some of those key characteristics of insights according to the literature.

At the outset of the experiment, there were initial subjective mood ratings that every participant had to complete. After each trial, there were retrospective questions about the experience of working on the problem, to check for possible prior exposure to the problem, to see how discontinuous or continuous the problem solving experience was, to see how sudden the solution was arisen, and to understand the mood of a participant while s/he was working on the problem (Appendix II contains the exact rating questions). These retrospective subjective ratings served as a benchmark for comparison with our physiological measures, as well as to validate the insight vs. non-insight nature of our particular problems. By utilizing single-step puzzle problems, routine problems, and subjective ratings on tasks, we are able to make the following specific hypotheses.

1.4 SPECIFIC HYPOTHESES

Figure 4 illustrates the hypothetical relationships between task nature, task experience (mood, problem solving experience and solution perception), and physiological measures. We propose that the nature of the task is going to be reflected in different patterns in physiological changes (relationship III in Figure 6, as in our first specific hypothesis). We also expect that the two kinds of tasks are going to result in different subjective ratings of task experience and mood (relationship I in Figure 6, as in our second specific hypothesis). We propose that some of these task experiences and moods during the tasks will serve as predictors for the patterns of physiological changes resulted (relationship II in Figure 6). Lastly, we propose that certain variations of task experiences and moods are mediators for relationship III, explaining why differences in task nature between puzzle problems and routine problems result in different patterns in physiological measures.

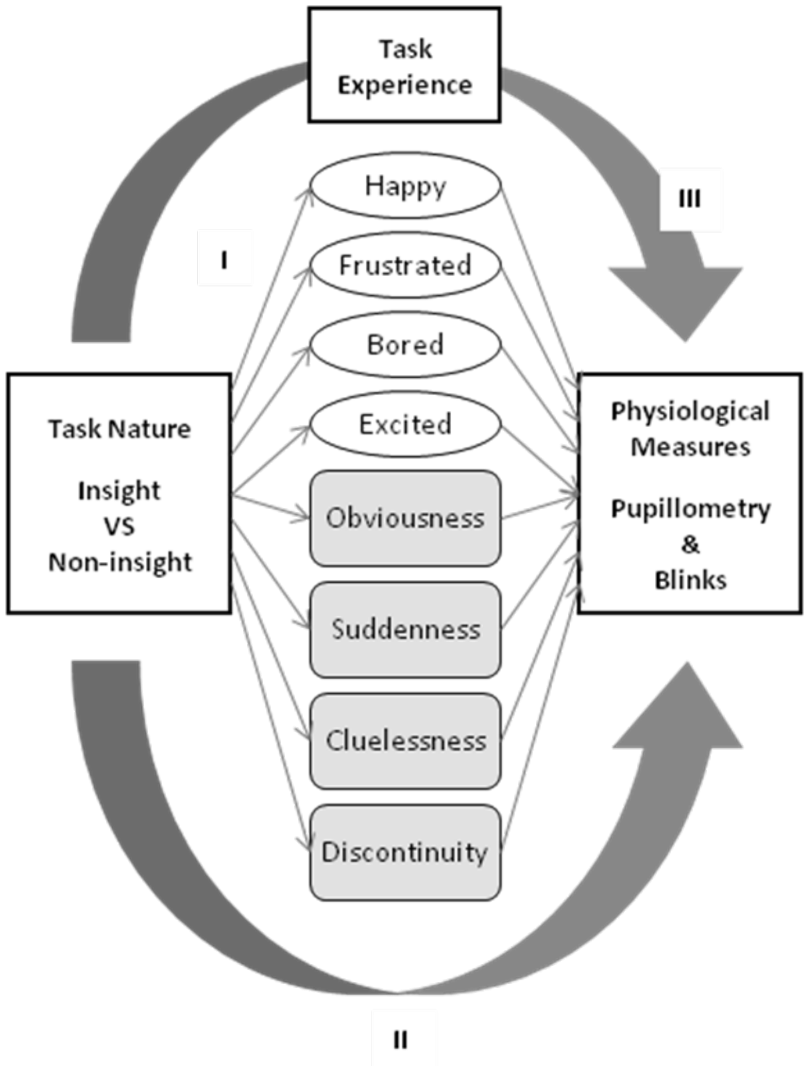


Figure 6. Hypothetical relationships between task nature, task experience and physiological measures.

1.4.1 Physiological measures between tasks

1.4.1.1 Puzzle problems

Based on our puzzle problems being harder and having a more sudden emergence of solutions because they are adopted from the insight problem solving literature, we have the following hypotheses about the pupillary and blink profiles of problem solvers around the time of 'Aha!' solutions:

Working memory release and blinks

The solution search for puzzle problems likely requires more cognitive resources, not only just prior to the moment of solution, but throughout the solution process. For example, it requires ongoing resources for monitoring (meta-cognition) of progress in the solution search, such as what the heuristic school of thought of insight problem solving has suggested (Kaplan & Simon, 1990; MacGregor et al., 2001). By contrast, we utilized incremental tasks that can readily be divided into steps and the search path for a solution is clearly defined for each problem at the very beginning. The routine problem has little need for the problem solver to keep track of the alternative heuristics that they have tried out.

Thus a critical difference between these solution searches is: one kind allows incremental freeing up of working memory, and the other kind does not. As a result, we predict sustained difference in the levels of blink frequency between the two kinds of problem because blinks track the release of cognitive resources. The two kinds of problems should return to a similar blink level after solutions have been reached possibly within a few seconds.

Effort in trying to interpret a novel answer on pupil dilation

We expect their pupillary profiles in the data to indicate that processing load between routine problems and puzzle problems are similar until the time when insights during puzzle problems are about to emerge (although, it is also possible that problem solvers may have discovered the solution before the actual execution of button press).

We are expecting the pupillary profile will have a peak at the solution emergence, and the peak will be higher among puzzle problems than among routine problems (see the pink line versus dark red line at time point zero and slightly after time point zero). That effect occurs because the pupil dilation activities may represent a cognitive effort of interpreting a novel answer and trying to integrate the novel answer into the problem solver's existing knowledge as Gruber (1995) have suggested. This effect does not require an absolute novelty of solution. Indeed, the degree of novelty of a solution may vary. For example, solutions of some difficult insight problems may not have been thought of at all while some solutions are only uncommon, so that they do not come into problem solvers' mind immediately. In either case, as long as a solution is uncommon, it already has some novelty that will trigger larger pupil dilations.

Pupil and arousal levels following solutions

We expect that problem solvers will feel more excited after solving a puzzle problem than after solving a task that requires incremental steps. Janisse's review (1976) suggests that the emergence of a novel answer can produce an emotional reaction to the solution, in addition to reflecting the cognitive and emotional processes that produced the answer. The arousal that a problem solver experiences during the sudden emergence of a novel solution is a quite widely used characterization of the subjective 'Aha!' experience.

Transforming the feelings of arousal into pupillary profile, we are expecting larger pupil dilations among puzzle problems than among routine problems because of a higher arousal level during discovery of novel solution. (See the grey line versus black line in Figure 6b). Pupil dilations before the solution are more due to cognitive load of producing a novel solution, whereas great pupil dilation after the solution are likely due to the 'Surprise' emotion itself from the discovery of a novel solution.

For hinted trials versus trials that were not hinted, we did not expect large differences between their physiological measures. However, we expected that participants would experience problem solving slightly differently because of the hinting, and these differences will be examined.

1.4.1.2 Routine problems

Effects of working memory load on routine problems

We expect the pupillary response and eyeblink responses of our participants on the routine problems to be mainly related to working memory load. In every routine problem in our experiment, a participant only had to fulfil definite and well-defined problem goals.

As we will describe greater detail in the next section, our routine problems only involve simple arithmetic and counting. We think that only having two simple, incremental goals in the routine problems demands less effort, and there is only potentially one transformation substitution operation to complete the task, which also does not lead to a high demand in working memory load (Anderson, Reder, & Lebiere, 1996). We expect that pupillary profile will be relatively gentle, and if there is a peak, it would be at the end of the counting, due to completion of solution search. Eyeblinks are expected to be of higher relative frequency throughout the whole task relative to puzzle problems. By contrast, interpretation and solution search for our puzzle problems requires the simultaneous creation of an imagined operation, examining each solution move, and also a memory stack to hold all the solution moves that a participant had already examined.

1.4.2 Effects of responding on blinks and pupils in both tasks

In both the puzzle problems and routine problems, we expect that there will be surges of blinks at the exact time of answering due to motor movements to press a key. There will also be peaks of pupillary dilations at the time of key press due to motor preparation and execution.First paragraph.

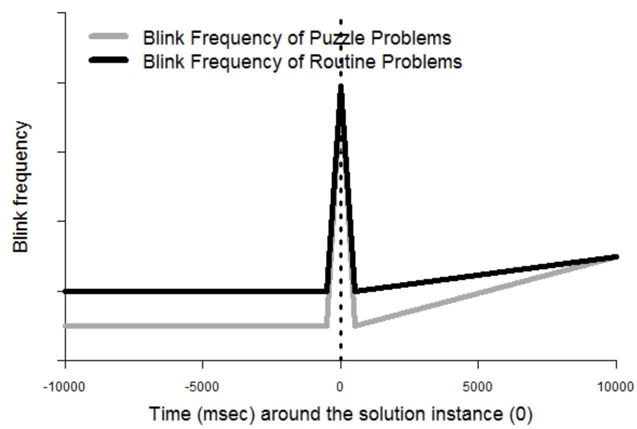


Figure 7. Prediction of blink frequencies.

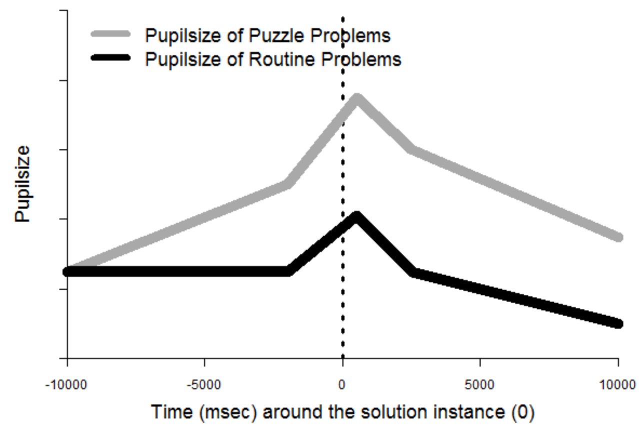


Figure 8. Prediction of pupillary responses.

Figure 7 and Figure 8 are predictions about the temporal progressions of pupillary changes and blink frequencies based on the above hypotheses as well as shared predictions across the three existing theoretical approaches of insight problem solving. Although not explicitly stated and due to different factors, all three existing theoretical approaches have a shared implication that problem solvers should have a higher load and more aroused state when dealing with puzzle problems in the first half of the pupillary progression curves before their solution time. The existing theoretical approaches also imply that there should be differences in the blink frequencies around the time when the solution emerges. On top of that, we propose that solutions from puzzle problems need more cognitive resources to process after being reached, as Gruber (1995) proposed that cognitive gap-filling, and structural transformation during discoveries are demanding processes of their own, leading to have a longer tail of blinks curve and a gentler slope of the pupillary response curve after the solution before both curves of blinks and pupillary changes eventually converge to their routine problem counterparts.

1.4.3 Relationship between mental events, blinks, and pupillary responses

We also want to explore the temporal relationship between pupillary changes and variations of blink frequency in a similar manner as in the study by Siegle et al. (2008). The button press events in our design can be thought of as internally driven by the participant when a solution is reached, and this can be different from the event of stimulus onset, which was the focus of temporal investigations in their study. We speculate that there is likely to be some resemblance between externally driven mental events and internally driven mental events, however, it is possible there will be a difference in manifestations of these two kinds of events on pupillary responses and blinks.

In the study of Siegle et al. (2008), by superimposing z-scored relative blink frequencies and z-scored pupillary responses, it was revealed that blinks tend to consistently precede pupillary responses during the onset of stimuli for digit-sorting and Stroop tasks, before pupillary responses catch up the lag and exceed the degree of blinks. Comparing our design with theirs, their participants were not likely engaged in mental processing of the tasks before the stimulus onset, while our participants should already be engaged in some mental processing of solution search well before solution time. However, assuming there is no substantial difference between internally-driven mental events and external-driven mental events, and considering the characteristics of discovery-based puzzle problems versus the characteristics of instruction-based routine problems, we expect that there should be some difference in the temporal relationship between the two kinds of tasks. We propose that a pattern of blinks preceding pupillary dilations should occur close to the solution time for puzzle problems, and that pattern of blink-precedence should occur much longer before the solution time for routine problems – approximately right after the onset of stimuli when a participant start processing the stimulus. In other words, the blink-precedence pattern would describe the flash of insight for puzzle problems, while it would describe the onset of mental manipulations based on instruction for the routine problem. Figure 9 and Figure 10 are illustrations of our proposed difference between the temporal relationships between pupillary responses and blinks profiles of puzzle problem versus routine problems.

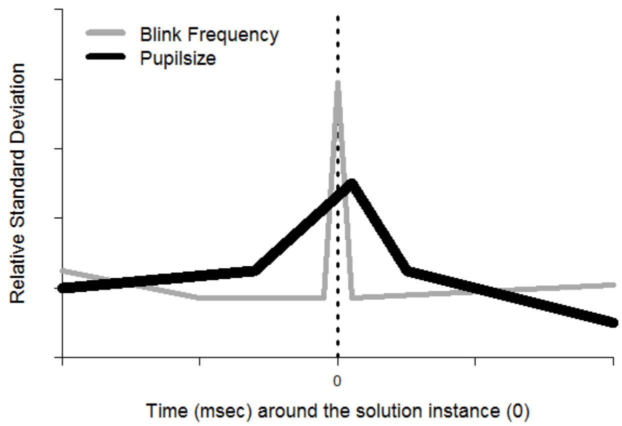


Figure 9. Prediction of temporal relationship between pupillary responses and blinks in puzzle problems in z-score units.

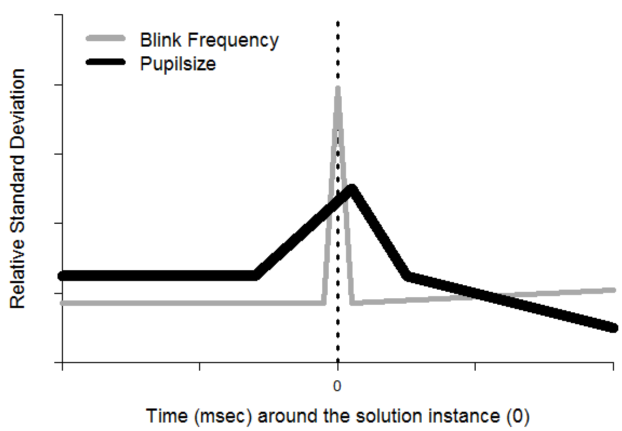


Figure 10. Prediction of temporal relationship between pupillary responses and blinks in routine problems in z-score units.

1.4.4 Subjective ratings between tasks

We expect that the subjective ratings between the puzzle problems and the routine problems be different. In particular, as a manipulation check, we expect the puzzle problems will be rated higher in “Cluelessness”, “Discontinuity”, “Obviousness”, and “Suddenness” than the routine problems. We also expect our participants to feel more “Frustrated”, more “Happy”, more “Excited” and less “Bored” during the solving of the puzzle problems than during the routine problems.

1.4.5 Relationships among task nature, subjective ratings, and physiological measures

Apart from predicted differences in physiological changes and subjective ratings between puzzle problem tasks and instructed routine problems, we are also interested in the overall relationships among these measures. In the study by Metcalfe and Wiebe (1987), subjective ratings during insight problem tasks were found to be poor predictors for solution time and success, and reliable only for incremental problems and algebra problems. Contrary to their synchronous subjective reports of “warmth ratings” and “feelings of knowing”, we explore whether our retrospective subjective ratings and synchronous physiological measures of pupillary responses and blinks can describe the difference between our puzzle problems, and our routine problems, as a manipulation check of the insight problem vs. noninsight problem manipulation as well as to provide more information regarding the phenomenological different between conditions that corresponds with blink and pupil effects. The relationship is denoted by arrow III in figure 6.

2.0 METHOD

2.1 PARTICIPANTS

Forty-five undergraduates of University of Pittsburgh enrolled in an Introductory Psychology class participated in the study as part of their course requirement (27 females, participant ages ranging from 17 to 30). No participants had any significant visual and hearing impairment. We also verified that they do not have prior experience with similar puzzle problems.

2.2 TASK

Both puzzle problems and routine problems in our experiment made use of Roman numerals and the concept of “matchsticks” making up a Roman numeral as in the study of Knoblich et al. (1999). In every puzzle problem and routine problem, a matchstick arithmetic expression is presented to the participant. A matchstick arithmetic expression is an arithmetic statement consisting of Roman numerals and a combination of arithmetic operators (‘+’, ‘–’, ‘=’), with participants imagining that every numeral and operator is made up of matchsticks.

2.2.1 Puzzle problems

In the puzzle problem version, the matchstick arithmetic expressions presented to the participants are false arithmetic statements. The goal of these matchstick arithmetic puzzle problems is to correct the initially false arithmetic statement by moving a single matchstick. Although participants may invent many possible moves of these matchsticks, there are rules that participants have to observe:

(i) they can only move a single matchstick;

(ii) a matchstick cannot be discarded or hidden in the expression;

(iii) the result must be a correct arithmetic expression, and one that we can evaluate for correctness, i.e. an expression without an equal sign cannot be evaluated;

(iv) a slanted matchstick ‘*I*’ cannot be interpreted as ‘I’, and ‘V’ and ‘X’ must consist of two slanted matchsticks;

(v) no other varied form of equal sign can be used, i.e. signs such as unequal≠’, more -than ‘>’, more-than-or-equal-to ≥’, less -than ‘<’, and less-than-or-equal-to ≤’ are forbidden;

(vi) there can be no multiplication or division in the corrected arithmetic expression;

(vii) there can be no negative numbers in the corrected arithmetic expression.

Before the experiment, every participant was presented a simpler problem (Figure 11), as an example for them to try – the solution to problem Figure 11 is to move one matchstick from a ‘II’ on the left hand side to the ‘II’ on the other side of the equal sign, so that the new expression will read “II + I = III”. Each of the matchstick arithmetic puzzle problems utilized in this experiment consists of three roman numbers with two arithmetic operators between them. All of them can be solved with a just single move. Figure 12 to figure 15 are other examples of puzzle problems that we used during the experiment. Readers are recommended to try solving each of these problems before reading the more detailed descriptions of them that describe solution elements. All problems and their solutions are listed in Appendix I.

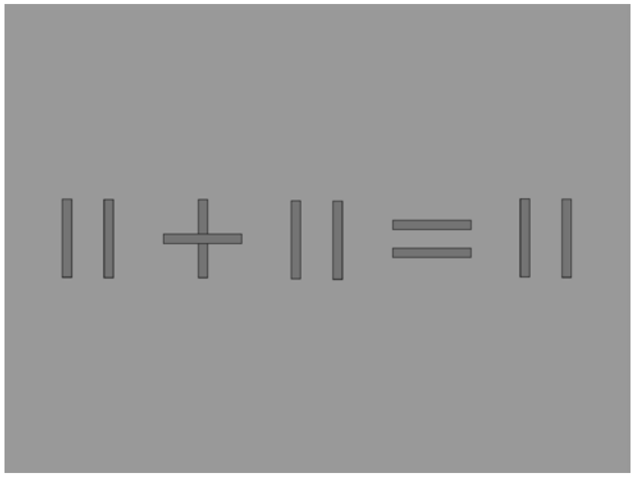


Figure 11. The example matchstick arithmetic puzzle problem.

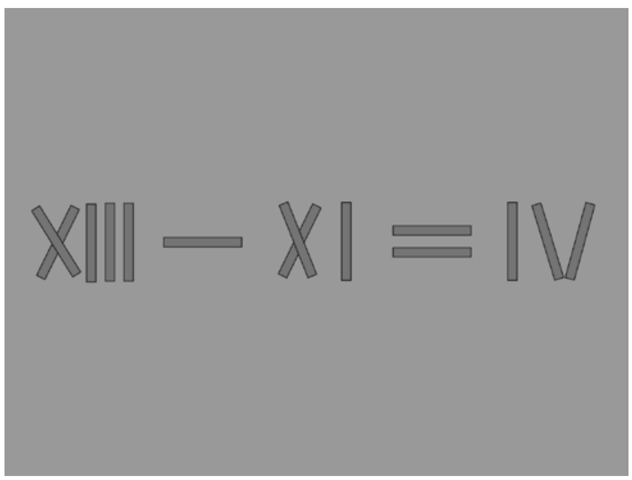


Figure 12. A matchstick arithmetic puzzle problem of type I.

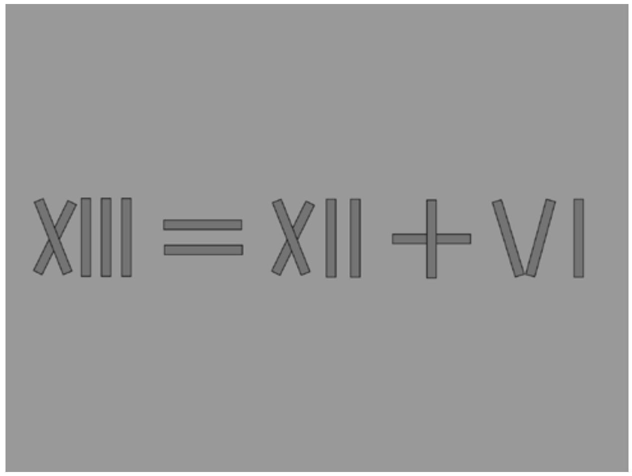


Figure 13. A matchstick arithmetic puzzle problem of type II.

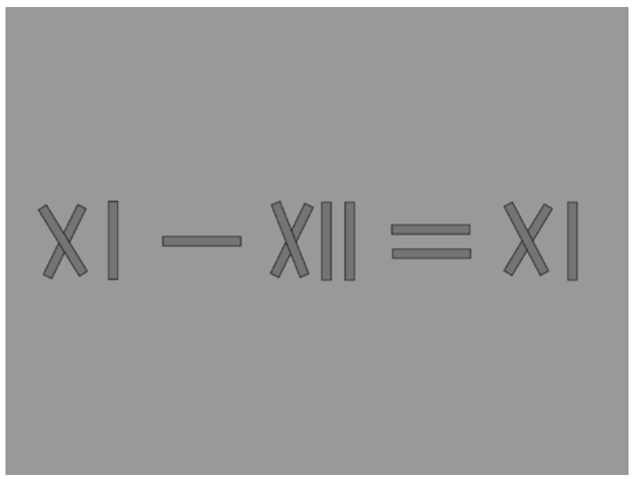


Figure 14. A matchstick arithmetic puzzle problem of type III.

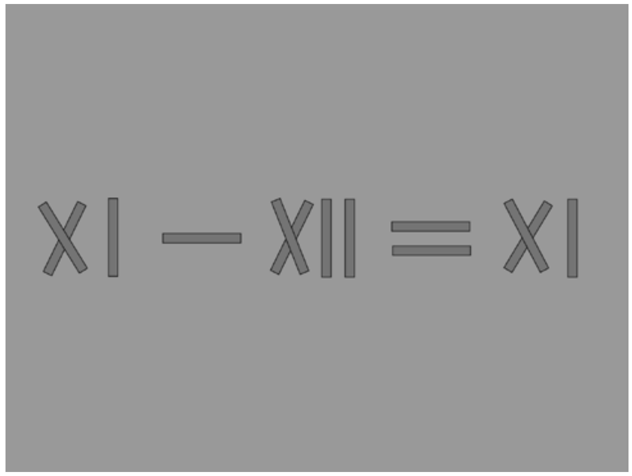


Figure 15. A matchstick arithmetic puzzle problem of type IV.

Similar to the study of Knoblich et al. (1999), difficulty level of actual problems varied systematically according to the constraint that a problem solver has to relax. Problems that require the simplest and most intuitive kind of moves between numeral(s) and/or arithmetic operator(s) by relaxing the easiest kinds of constraints (e.g. value constraint and/or operator constraint) are classified as type I; problems that may be affected by inhibition of return are classified as type II; problems that require decomposition of a tighter chunk are classified as type III; problems that require relaxation of a stubborn constraint (e.g. tautology constraint) are classified as type IV. See Knoblich et al. for additional details regarding these constraints.

2.2.2 Routine problems

We want to have our routine problems closely resemble our puzzle problems on many dimensions and yet having a contrast with respect to the insight nature of the problems. Specifically, we have three concerns in the design. First, we want participants to perceive these

Roman numerals as numerals and perform similar mental numeric operations and evaluations of Roman numeral expressions. Second, we also want them to break down each Roman numeral and perceive the matchsticks that make up each numeral. Third, we want a computational task that is routine in nature but involves a basic memory load as is found in most problem solving tasks. As a result, we designed our routine problems as having an appearance of simple arithmetic addition and subtraction problems expressed in Roman numerals:

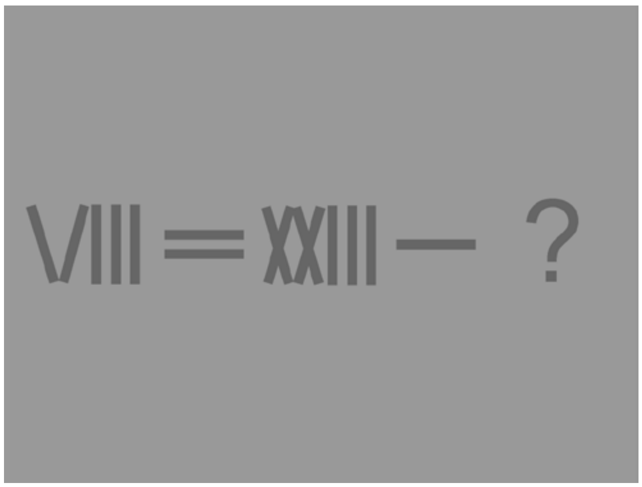


Figure 16. An example of a routine problem.

In addition to completing the arithmetic task, i.e. to figure out the Roman numeral in the question mark location, participants were requested to break down the corresponding Roman numeral and then to count the number of match sticks in the missing numeral. To increase the memory load to less trivial levels for a problem solving task, participants were requested to count not just the matchsticks in the missing numeral, but also the number of match sticks in the operators. That is, the whole task is to figure out what the missing numeral is, and then to add up its number of match sticks with the total number of match sticks in the arithmetic operators. Deliberately asking them to count some of the matchsticks in the routine problem prevented our participants from perceiving the expressions as just Roman numerals and operators. Also, the operator counting task increases the complexity of our routine problems so that they can more closely match our puzzle problems in overall difficulty. In the above example, the missing numeral is 'XV', which has 4 match sticks. And hence, the final number that the participant has to report is 4 plus the number of match sticks in the operators ('=' and '–'), i.e. 4 + 3, which is 7. The overall trial structure in a session is illustrated in Figure 17:

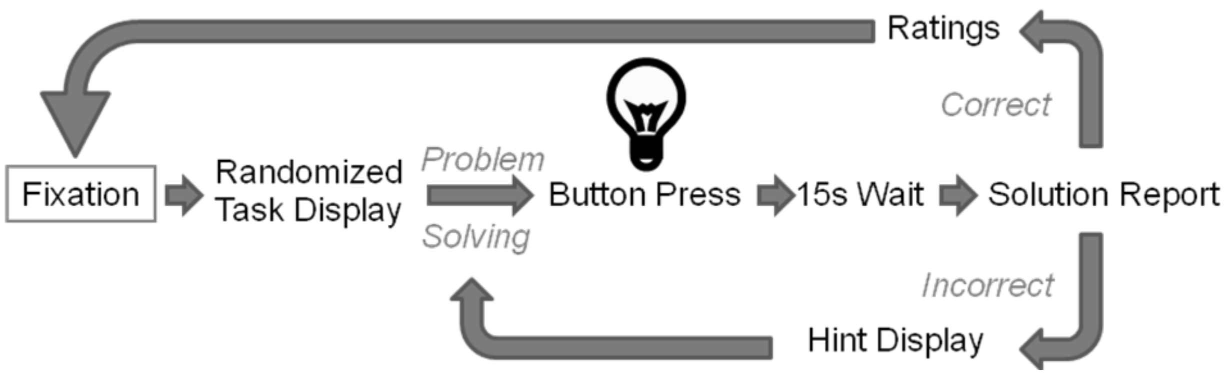


Figure 17. Trial structure for both insight and non-insight problems.

2.2.3 Subjective ratings

To explore the mood and the subjective perceptions during problem solving, we asked participants to retrospectively report their mood and their perceptions of the problem, divided into two sets of ratings to be completed after they finished each task. In the first set, there are four ratings concerning their mood along the dimensions of being "Happy", "Frustrated", Bored", and "Excited" about the task that they just completed. A second set of ratings involved their experience of the problem solving process and how the solution finally emerged. There were four ratings in the second set: "Cluelessness" about the solution search, "Discontinuity" about the solution process, “Obviousness” about the solution when it emerges, and “Suddenness” about the emergence of the solution.

2.3 APPARATUS

2.3.1 Room and stimuli presentation

Stimuli were presented to the participants on a 17-inch LCD computer screen using the E-Prime® 2.0 package of Psychology Software Tools, Inc. The vertical height of stimuli on-screen was approximately 7.5 cm, subtending an approximately 7 degree visual angle for the participants. Stimuli and on-screen instructions were all coloured in greyscales, so that during the change of stimuli within a task, the degree of change in illumination would not induce significant pupil size changes. The illumination of the room was kept constant.

2.3.2 Eye-tracking setup

Pupil size and blinks were monitored with a Tobii® eye-tracker model 1750. Distance between the eye tracker and the face of the participants was kept constant at 55cm by asking them to put their chin on the chin-rest fixed on the edge of the experiment bench. Eye data were sampled at a frequency of 60Hz on both eyes using an embedded infrared camera of the eye tracker through the software package developed by Tobii Technology.

2.4 PROCEDURE

The experiment involved two sessions spread across two different days to insure sustained levels of effort across the many puzzle problems and routine problems. Every participant completed their second session within 3 days of the first one. Each session consisted of eight puzzle problems and eight routine problems arranged in a random sequence. A 30-second break occurred after each problem. Participants solved 16 new problems but of similar kinds in the second session.

In order to make sure that our participants were all comfortable with Roman numerals, each experimental session began with training, consisting of presenting Roman numerals from one (I) to thirty-five (XXXV) and then giving a recognition task on these numerals for 5 minutes. Following the Roman numeral training, we described the two kinds of puzzle problems and routine problems, with example problems of each kind.

2.4.1 Initial subjective mood ratings

After the training, every participant answered a set of baseline mood related questions. There were four questions that they answered to indicate how they are feeling at that moment of the day: how happy, how frustrated, how excited, and how bored.

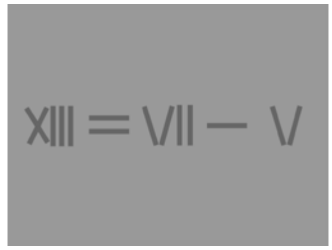
2.4.2 Solution attempt and Incremental hinting

2.4.2.1 Puzzle problems

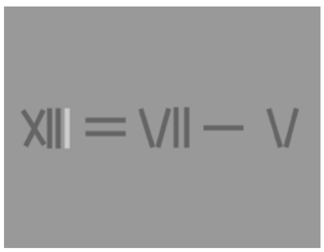
When a participant had a solution in mind, he/she was instructed to press the spacebar, and then they had to silently wait for another 15 seconds before giving out the answer verbally. This 15 seconds period served as a buffer to record their eye data before they started to perform any verbal or manual activities that might influence the eye data.

To insure that every participant had enough successful cases in solving these puzzle problems to produce sufficient statistical power in the analyses, a sequential series of hints were provided to participants. If the participant could not come up with a solution after the first sixty seconds, a first hint would appear: the match stick that should be moved will be highlighted (See figure 18b). After another 30 seconds, a second hint would appear: a square would appear to indicate the area to which that particular highlighted match stick should be moved to (See figure 18c). After a further 30 seconds, a third hint would appear: an arrow would appear to point out a more exact location and orientation that the particular match stick should be positioned (See figure 18d). The solution to the problem would be shown when a participant could not provide a solution after being shown the third hint for more than 30 seconds (See figure 18e).

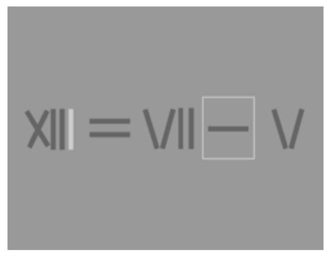
If the answer provided was not correct, the next hint was shown to the participant and he/she would continue to work on the same problem. For instance, if an attempted answer after the first hint was incorrect, the second hint would be shown, and the participant would continue to work on the problem; and if an attempted answer was incorrect after the third hint, he/she would be shown the solution.



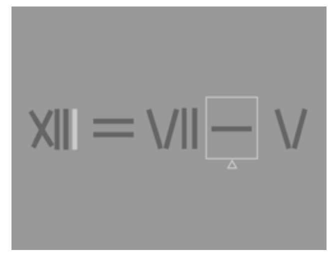
a. A match stick puzzle problem.



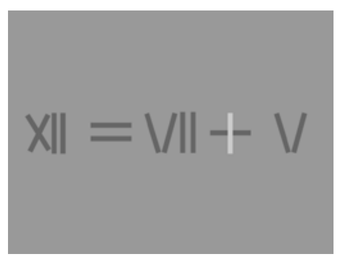
b. The first hint.



c. The second hint.



d. The third hint.



e. Solution.

Figure 18. Incremental hinting of a match-stick puzzle problem.

2.4.2.2 Routine problems

For routine problems, participant also had to press the spacebar to indicate that he/she has a complete solution in mind, and then they also had to silently wait for another 15 seconds before giving out the answer. They were given verbal feedback by the experimenter about the correctness of their answers. There was no incremental hinting for our routine problems because the format of these problems does not require incremental hinting.

2.4.3 Post-task subjective ratings

There were 9 retrospective questions about the experience of solving each problem participants has to answer after every problem. Four questions, similar to the baseline mood questions, were about being happy, frustrated, excited, and bored. Another four questions were about the retrospective cluelessness, discontinuity, obviousness, and suddenness about the problem solving experience and the subjective perception of solution. There was also 1 question to check if the participant has worked on the problem task before in somewhere else (See Appendix II for these questions).

2.5 DATA REDUCTION

2.5.1 Filtering of trials for analysis

For both the puzzle problem and routine problem trials, only trials with correct answers were included in the main analyses. Among 1424 trials we collected without equipment failure, 1213 trials (85%) were correctly answered. Also, trials that had no solution attempts during the trial duration were discarded, resulting in an elimination of 4 trials. Trials comprised of excessive blinks (over 55% of the 60Hz eye data sampling during a trial) were removed from the analyses of pupillary and blink responses as likely to due to equipment error or participants with temporary vision issues (resulting in the exclusion of 13 puzzle problem trials and 20 routine problem trials for the first session, 17 puzzle problem trials and 23 routine problem trials in the second session, a total of 73 trials across both sessions).

2.5.2 Calculation of blink frequency waveforms

We adopted a blink waveform operation similar to Siegle, Ichikawa and Steinhauer (2008) by defining blinks as changes in pupil dilation occurring too rapidly to be actual dilation or contraction. While we have ensured that the head positions of our participants were fixated in constant and detectable distance, we defined a data point as an eye blink when the size of pupil of a participant became so small that: 1) the system can only record one eye, and has no way of determining if that was the left or the right eye; 2) the system is fairly confident that the actual gaze data is missing; or 3) the actual gaze data is definitely missing. Data points of 3 or less in between two blinks were also identified as blinks because intervals between blinks of less than 0.05 seconds were unlikely to represent periods of clear vision. Both blinks and the interval between them were judged to be part of the same single blink.

Samples that satisfied blink criteria were coded as 1, and other samples were coded as 0. Then samples for each trial were interpolated to contain equal numbers of 1 mini-second data points. Average frequency of blinks at each data point for each condition (i.e., number of trials that blinks occurred at the data point divided by total number of valid trials in the condition) was then calculated to yield condition-related blink-frequency waveforms. The mean blink frequency of the first 10 samples preceding the onset of stimulus for each trial was then subtracted from the sample to obtain relative blink frequency.

2.5.3 Calculation of pupillary waveforms

Pupillary data were reduced using a methodology similar to that of Siegle et al. (2008) and Granholm et al. (Granholm, Asarnow, Sarkin, & Dykes, 1996). For each trial, we used the average of the first 10 pre-task data points of pupil diameter reading to be the baseline of pupillary responses for that trial. In addition to removing trials that were comprised of over 55% blink moments, we performed linear interpolations to replace pupil size during the blinks throughout the data set using four samples before and nine samples after a blink. Siegle et al. (2008) suggest that this technique prevents interpolation to poorly estimate pre- and post blink pupillary measures because of partial eyelid closures.

2.5.4 Determination of windows of significant differences

Contrasts on pupil dilation were examined via statistical tests at each bin along the pupil dilation and blink waveforms. To control type I error for this large number of tests and given the temporal autocorrelation of the data, we used a 500 msec period as a window of significant difference, which is similar to the 35 samples at 60Hz obtained using Monte Carlo simulations on time series as reported in the study of Siegle et al. (2008). In addition, a stricter p-value of < 0.05 per sample was set as the threshold, instead of previously used < 0.1.

We used paired t-tests to examine within-subject effects of puzzle problems versus routine problems. We also wanted to investigate the relative patterns of blink and pupillary responses to see whether our study using more complex puzzle problems and routine problems replicated related results by Siegle et al. (2008). To do this direct blink to pupil size comparison, we compared the relative magnitude of blink frequency and pupillary response by rescaling data into standard-deviation units by dividing the relative change in millimeters or blink frequency from the pre-stimulus baseline values by the standard deviation of these changes over the entire waveform. We also used paired t-tests to compare the resulting waveforms in the same manner along the waveforms. Similar type I error correction methods as for condition-related comparisons were used, and statistically significant windows of differences were highlighted.

3.0 RESULTS

3.1 SOLUTION TIME AND RATINGS ACROSS TASKS

In both sessions, solution times of participants were significantly faster (session 1: $t(42)=-13.99$, $p < 0.001$, session 2: ($t(42)=-13.65$, $p < 0.001$) for the routine problems than for the puzzle problems, roughly 30 seconds vs. 60 seconds on session 1 and 25 seconds vs. 50 seconds in session 2. Thus, the two tasks were not entirely equated for task difficulty, although both tasks involved work distributed across roughly tens of seconds. In addition, participants took approximately 20% longer time to solve both problem types in session one than in session two (session 1: $t(41)=3.23$, $p < 0.005$, session 2: $t(40)=5.69$, $p < 0.001$).

For the puzzle problems, approximately half of the correct trials required some hinting, and relatively few trials requiring the extreme hints (See figure 19).

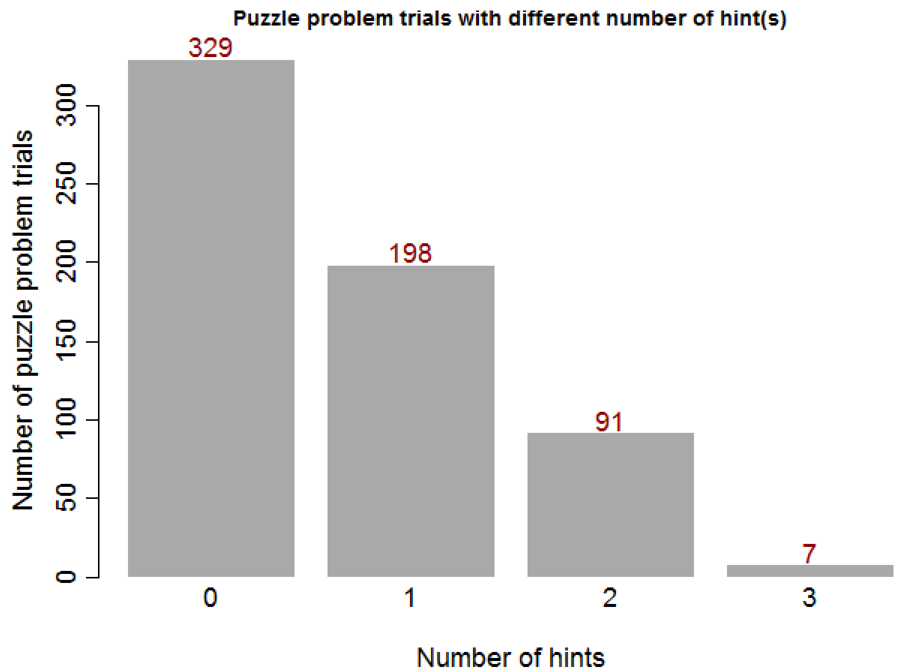


Figure 19. Histrogram of hinting for correct puzzle problem trials.

For mood ratings, participants felt happier about working on the routine problems than on the puzzle problems in session one. In session two, they felt more bored about the routine problems than on the puzzle problems. In both sessions, they felt more frustrated working on the puzzle problems than on the routine problems (See figure 20 and figure 21). Overall, the largest effects were for frustration levels, with large effects on session 1 and moderate effects on session 2. The other mood effects were inconsistent across days and were small effects when found.

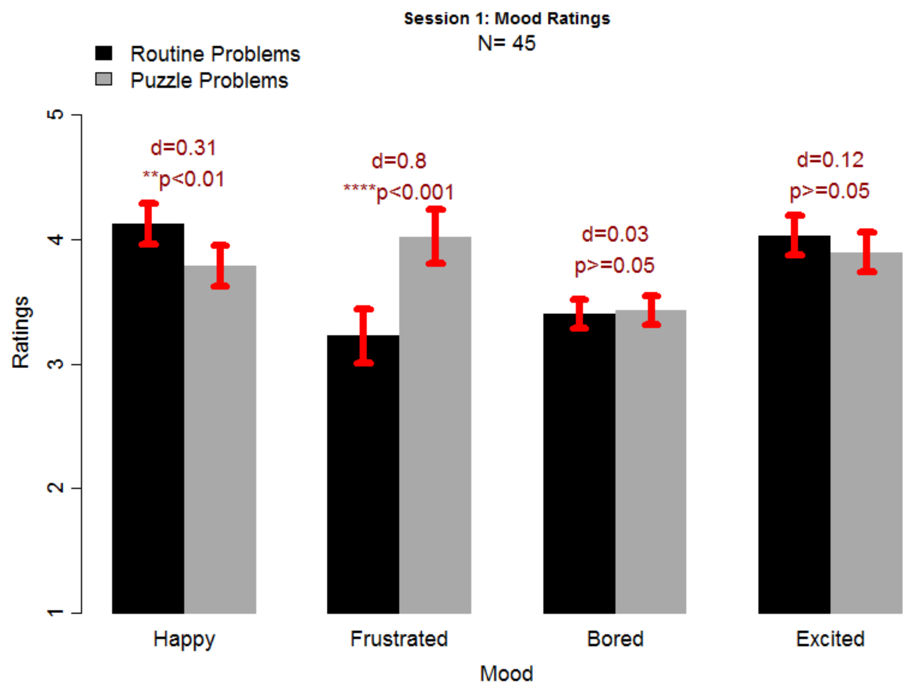


Figure 20. Mood ratings by insight vs. noninsight tasks in session one.

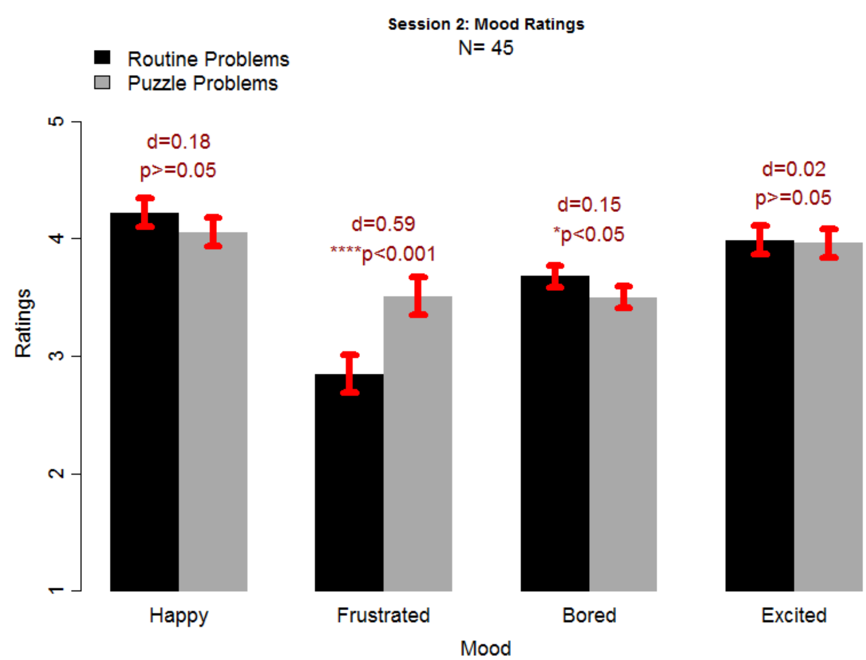


Figure 21. Mood ratings by insight vs. noninsight tasks in session two.

Mood ratings for puzzle problems that were hinted and that were not hinted are different. Participants felt less happy and considerably more frustrated when they solve the puzzle problems without hints. They also seems to feel slightly more bored but at the same time slightly more excited about puzzle problems that were hinted (See figure 22 and 23).

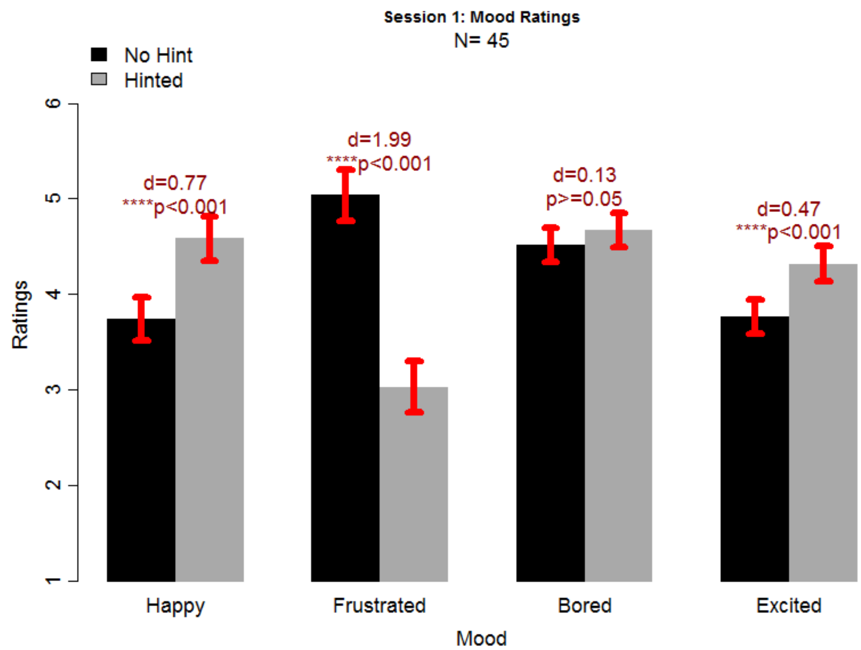


Figure 22. Mood ratings of hinted and nonhinted puzzle problems in session one.

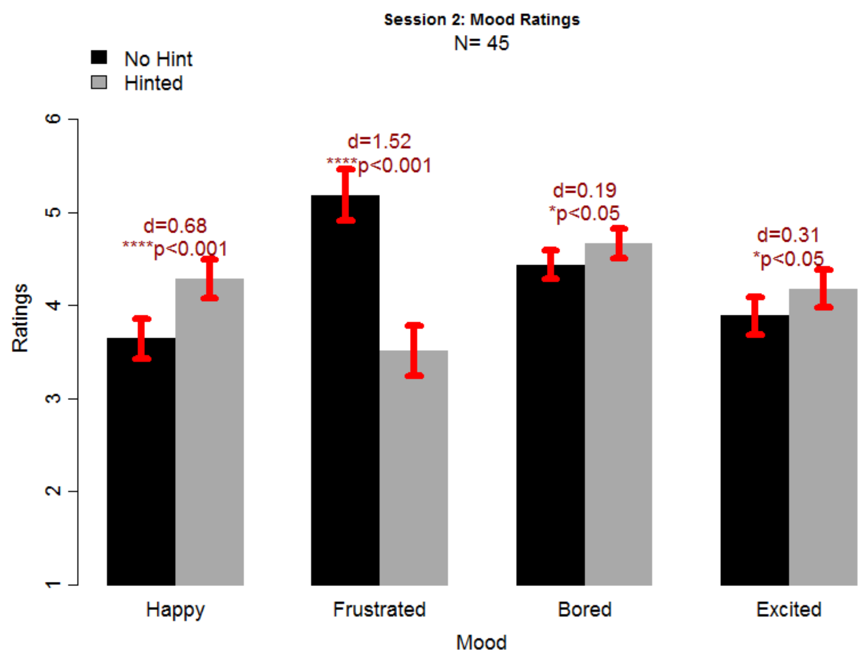


Figure 23. Mood ratings of hinted and nonhinted puzzle problems in session two.

For ratings about subjective problem solving experience and solution perception, participants felt the solution of the puzzle problems to be more obvious and more sudden than the routine problems in both sessions. They also felt the problem solving process of puzzle problems more clueless and more discontinuous than the routine problems in both sessions (See figure 24 and figure 25). These effects were all consistent across sessions, providing a strong manipulation validity check. The condition effects on the clueless and discontinuity dimensions were very large.

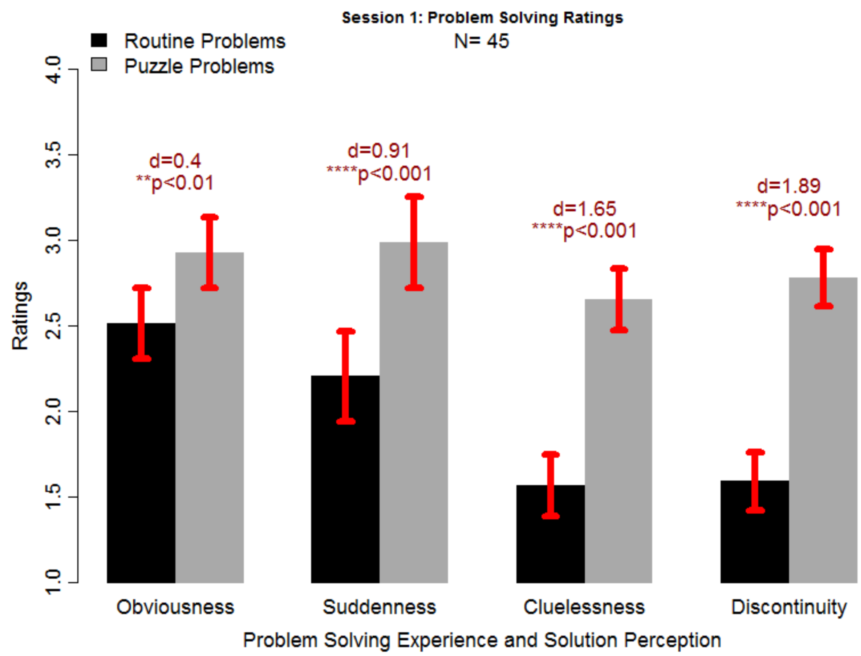


Figure 24. Problem solving experience and solution perception ratings in session one.

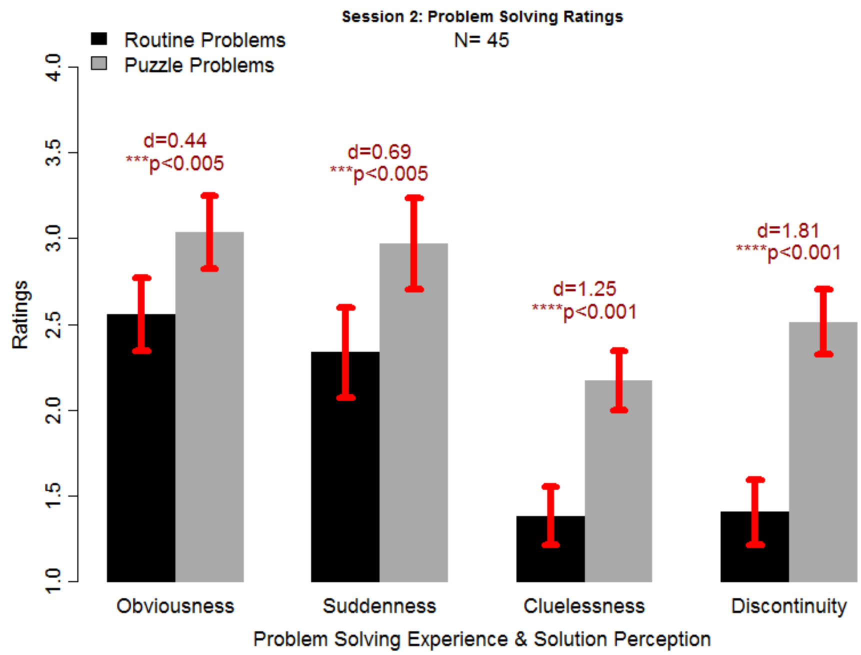


Figure 25. Problem solving experience and solution perception ratings in session two.

Problem solving experience ratings for puzzle problems that were hinted and that were not hinted are different. Participants felt higher in cluelessness and discontinuity in their experience of solving the puzzle problems when they were not hinted (See figure 26 and figure 27).

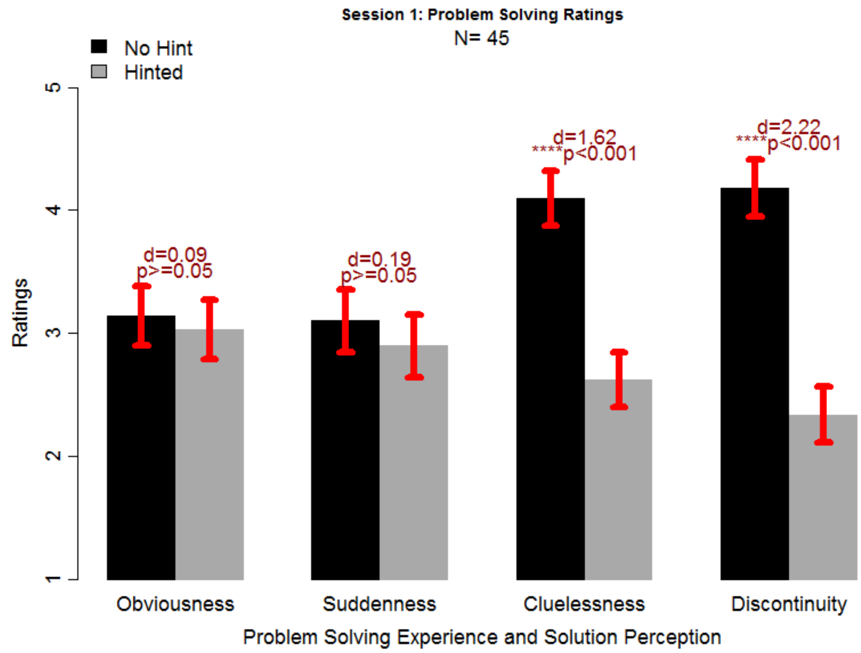


Figure 26. Problem solving experience and solution perception ratings of hinted and nonhinted puzzle problems in session one.

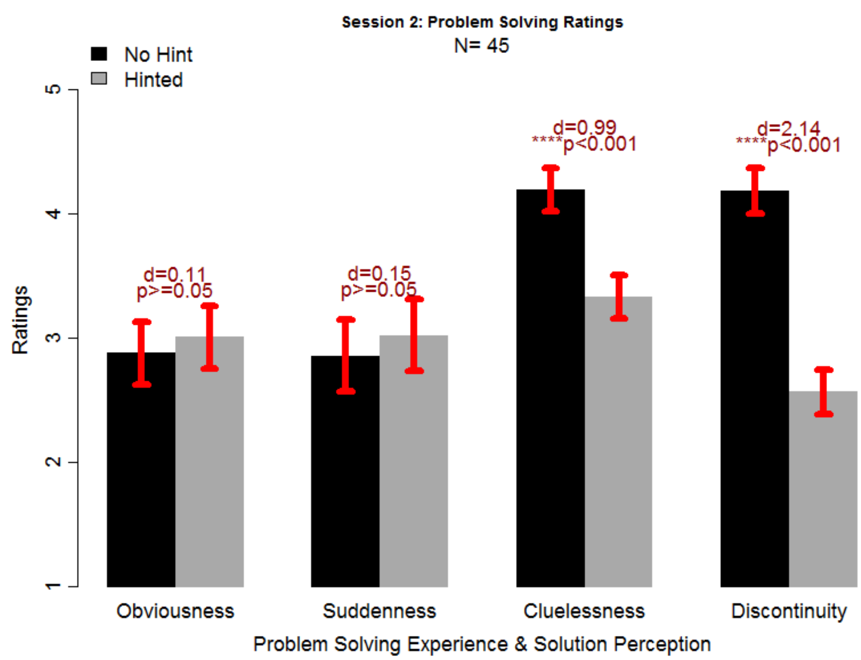


Figure 27. Problem solving experience and solution perception ratings of hinted and nonhinted puzzle problems in session two.

In order to make sure that differences in subjective ratings were a function of the particular type of task and NOT because of the difference in their solution times, we conducted addition correlation tests. By performing Spearman's rho correlation tests between solution times of routine problems and their subjective ratings, we found none of the ratings have a significant p-value in their Spearman's rho (See table 1). For the puzzle problems, we found the Spearman's rho correlations between solution times and the ratings of "Frustrated", "Obviousness" and "Discontinuity" are statistically significant (See table 2). The lack of correlations with time across tasks suggests that time per se differences could not explain the ratings differences across task conditions.

Table 1. Spearman's rho correlations between solution time of routine problems and corresponding ratings.

Subjective ratings	Spearman's rho correlations between solution time	
	Session 1	Session 2
Happy	-0.09	0.24
Frustrated	-0.17	-0.15
Bored	-0.01	-0.21
Excited	0.09	0.18
Obviousness	-0.19	-0.22
Suddenness	0.04	0.19
Cluelessness	0.10	0.06
Discontinuity	0.22	0.12

Table 2. Spearman's rho correlations between solution time of puzzle problems and corresponding ratings.

Subjective ratings	Spearman's rho correlations between solution time	
	Session 1	Session 2
Happy	-0.09	-0.22
Frustrated	0.23	0.35*
Bored	-0.12	0.11
Excited	-0.22	-0.18
Obviousness	-0.39**	-0.41**
Suddenness	-0.26	-0.13
Cluelessness	0.26	-0.01
Discontinuity	0.36*	0.01

Note: * p < .05, ** p < .01

3.2 PHYSIOLOGICAL MEASURES ACROSS TASKS

Figures 28a and 28b show the blink and the pupil proportion waveforms during session 1 for both routine problems and puzzle problems. Both blinks and pupillary response showed variations related to solution events at time 0. For blinks, statistically significant condition-related outbursts occurred starting from 8.19s before the solution button press and ending at 6.06s after the button press, with participants having more blinks during the routine problems than during the puzzle problem in that window ($t(42) = 2.02$, $p < .05$, $d = 0.62$). Pupillary responses of participants displayed statistically significant larger pupil dilations in the puzzle problems than in the routine problems starting from 4.46s before button press to 1.7s after button press ($t(42) = 2.02$, $p < .05$, $d = 0.62$).

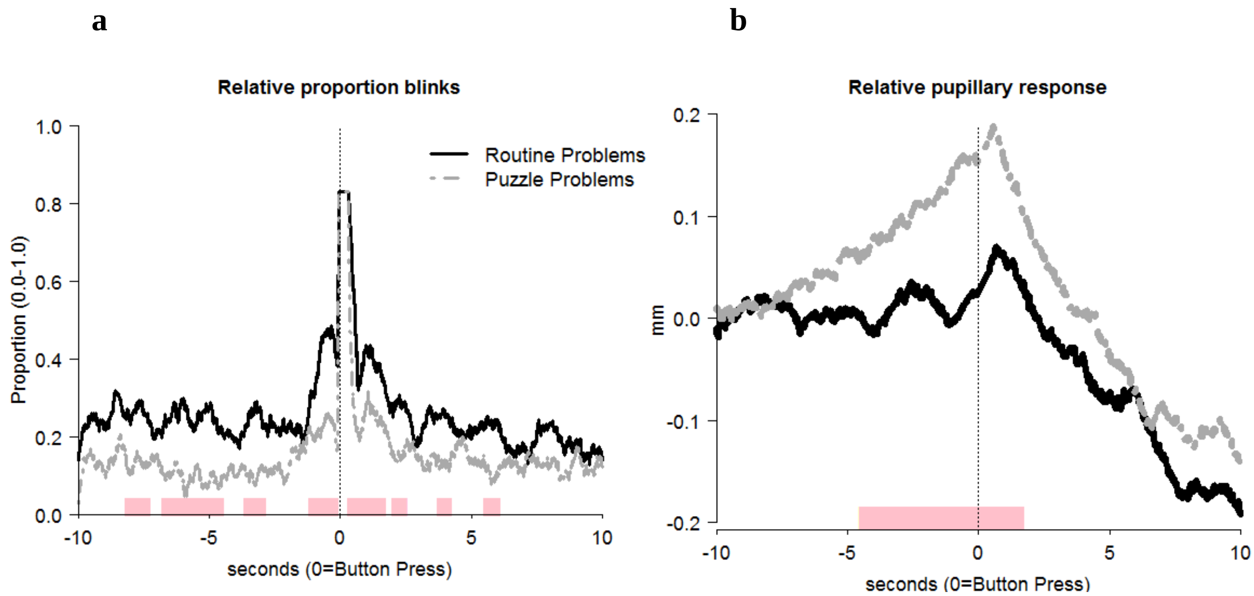


Figure 28. Blink frequency and pupil dilation around the solution time (0) during routine problems and puzzle problems in session one (n=43). Panel a shows the proportion of blinks for each task around solution time. Panel b shows pupil dilations for each task around solution time. In both panels, significant condition-related differences are highlighted above the x-axis ($p < .05$).

In session 2, pupillary response and blinks also showed variations related to solution events at time 0. Figures 29a and 29b show the pupil and the blink proportion waveforms during session 2 for the two kinds of tasks. As in session 1, participants have significantly higher intermittent outbursts of blinks in the routine problems ($t(42) = 2.02$, $p < .05$, $d = 0.62$), but this time in the window from 5.79s before the button press to 1.83s after the button press. That is, this analysis found fewer significantly different blink blocks between task conditions in session 2. The pupillary responses also displayed significantly more intense pupil dilations in the puzzle problems than in the routine problems in session 2, this time starting from 4.94s before the button press to 7.07s after the button press ($t(42) = 2.02$, $p < .05$, $d = 0.62$).

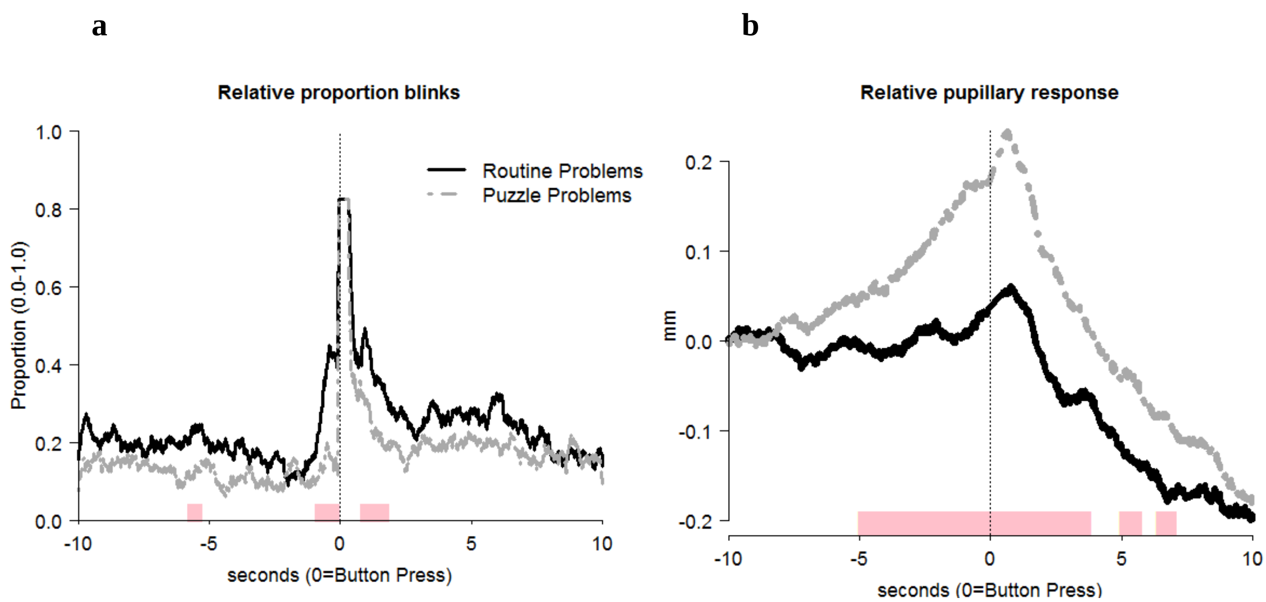


Figure 29. Blink frequency and pupil dilation around the solution time (0) during routine problems and puzzle problems in session two (n=43). Panel a shows the proportion of blinks for each task around solution time. Panel b shows pupil dilations for each task around solution time. In both panels, significant condition-related differences are highlighted above the x-axis (p <.05).

In both sessions (see Figure 28b and Figure 29b), we found that pupillary responses went well below baseline at 10s after the button press of solution declarations. At this 10s time point after the button press, participants had finished the task and were simply rehearsing the solution in their mind in order to report the solution at 15s after their button press. They were likely more relaxed than when they were at the baseline time point where they may have been vigilant in preparation of the trial or mildly anxious or curious to determine whether the next problem is a routine or puzzle problem.

Within puzzle problems, hinting and no hint trials were slightly different in blink and pupil profiles in session 1 (see figure 30a and 30b), but there were no statistically significant differences in either profile in session 2 (see figure 31a and 31b).

a

b

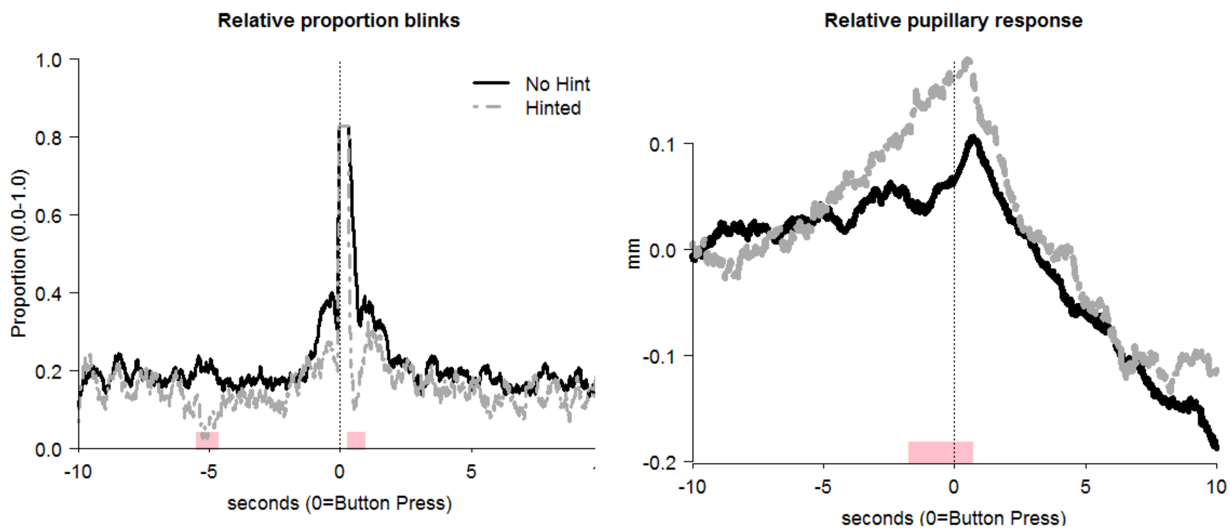


Figure 30. Blink frequency and pupil dilation around the solution time (0) during hinted puzzle problems and non-hinted puzzle problems in session one (n=43). Panel a shows the proportion of blinks for each task around solution time. Panel b shows pupil dilations for each task around solution time. In both panels, significant condition-related differences are highlighted above the x-axis (p <.05).

a **b**

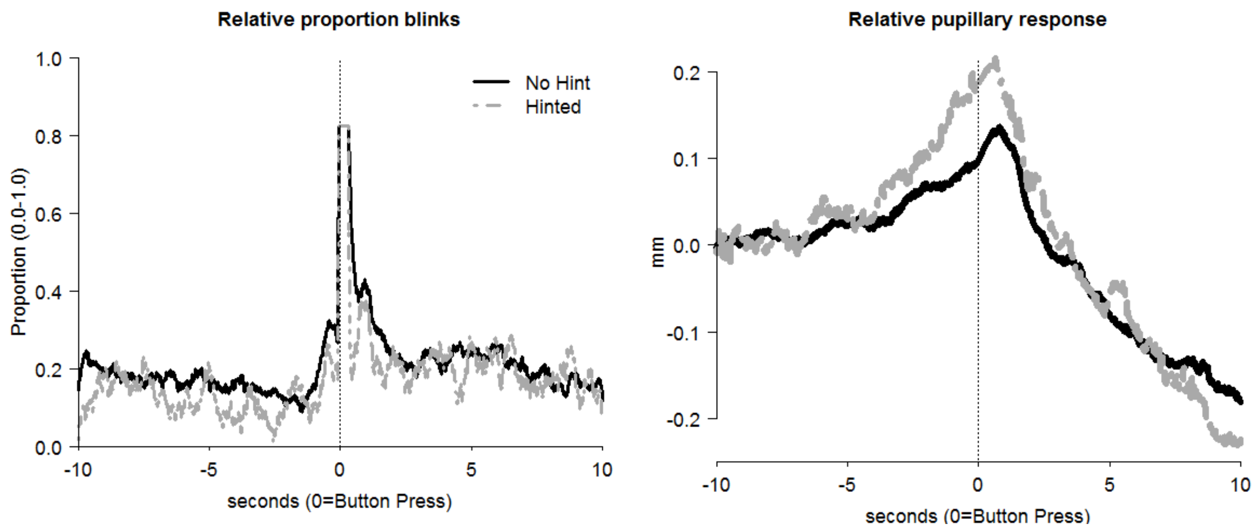


Figure 31. Blink frequency and pupil dilation around the solution time (0) during hinted puzzle problems and non-hinted puzzle problems in session two (n=43). Panel a shows the proportion of blinks for each task around solution time. Panel b shows pupil dilations for each task around solution time. In both panels, there was no significant condition-related differences ($p > .05$).

We also performed comparisons for the same kind of tasks across sessions. For routine problems, we found no significant difference between the waveforms of pupillary responses in the 20 seconds time window around solution time; there were short durations of differences in blinks before the button press during the routine problems across sessions (a 611ms period at 2.24s before button press and a 602ms period at 1.22s before button press, $t(42) = 2.02$, $p < .05$, $d = 0.62$). For puzzle problems, we found no significant difference between the blink waveforms across the two sessions; there were a short duration of differences in pupillary response after button press (a 635ms period at 733ms after button press, $t(42) = 2.02$, $p < .05$, $d = 0.62$).

3.3 TIME COURSES OF PHYSIOLOGICAL MEASURES FOR EACH KIND OF TASKS

As waveforms within tasks for session 1 and session 2 are quite similar, we combined the waveforms across sessions according to the corresponding kinds of tasks to further look at the relationship between blinks and pupillary responses for the routine problems and the puzzle problems. Figure 32a and figure 32b show blinks and pupillary response z-scored and superimposed with significant differences between the two kinds of activities highlighted above the x-axis. These data suggest that there were constantly lower blink activities within the 10 seconds window before button press (before button press: 9.51s to 8.65s, 8.52s to 6.71s, 5.06s to 1.16s) for the routine problems ($t(43) = 2.02$, $p < .05$, $d = 0.62$). For the puzzle problems, significantly higher-than blinks pupillary activities occurs not until 6.31s before button press time ($t(43) = 2.02$, $p < .05$, $d = 0.62$). In both kinds of tasks, a burst of blinks occurred slightly before peak pupil dilation.

Apart from the qualitative difference between these two kinds of problems, we also tested the statistical interaction of the differences between their two blinks and pupillary response curves. The regions with significantly different absolute differences between the two curves (i.e., a pupil/blink by puzzle/routine interaction) are highlighted in blue under the x-axis. These data suggest that apart from the characteristic qualitative differences between the pupillary response versus blinks profiles of these two kinds of problems, there were also absolute quantitative differences if we compare their corresponding differences from within their own pupillary responses and blink curves: larger differences for routine problems before button press 8.52s to 7.80s and 0.96s to 0.09s ($t(43) = 2.02$, $p < .05$, $d = 0.62$).

a. Routine Problems **b. Puzzle Problems**

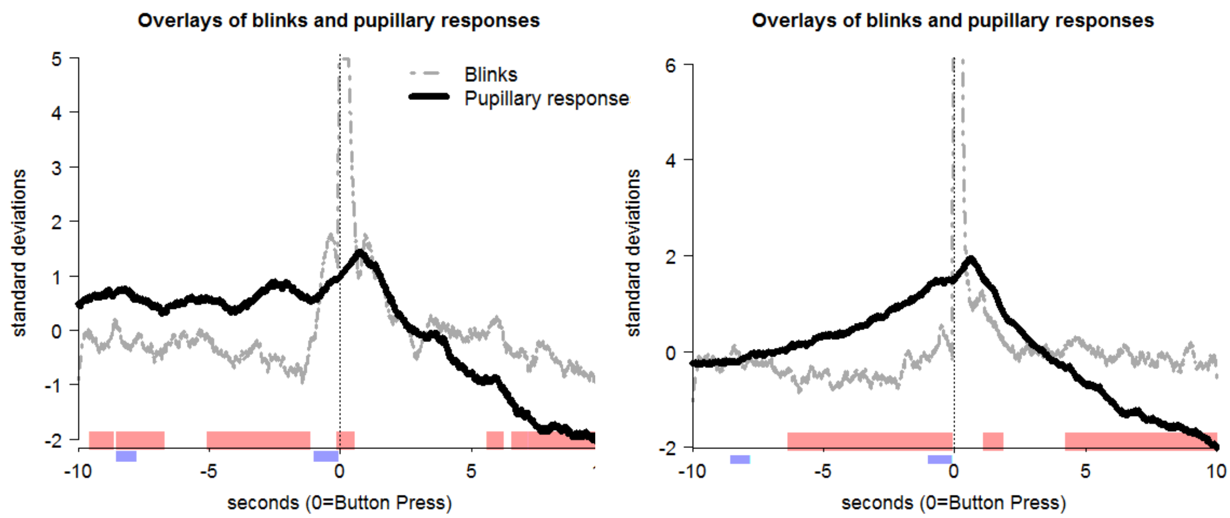


Figure 32. Z-scored overlays of the blink and pupil dilation data. Significant differences are highlighted above the x-axis ($p < .05$), and significant interactions between tasks and response modality are highlighted below the x-axis.

3.4 RATINGS AND PHYSIOLOGICAL MEASURES ACROSS TASKS

We used the information criterion procedure of Akaike (Akaike, 1974) to guide our regression analyses of the relationship between solution time and subjective ratings, and the overall relationship between physiological measures and subjective ratings with differences in task nature consider. The procedure measures the goodness-of-fit of statistical models to estimate relative measure of the information lost when a given model is used to describe the data in order to guide the usage of models.

Recall that solution times of routine problems were found to be unrelated to any of the subjective ratings. For puzzle problems, in a best-fit multiple regression model with

“Cluelessness”, “Discontinuity” and “Obviousness” as our predictors for solution time, we found that “Obviousness” significantly predicted solution time in puzzle problems (b = -0.43, t(37) = -2.89, p < .01) with the model explained a significant proportion of variance in solution time (adjusted R^2 = .14, F(3, 37) = 3.18, p < .05). In other words, when a participant feels more obvious about how to solve a puzzle problem, the shorter the solution time will be.

In order to examine effects of task assignment (routine problems versus puzzle problems) in our overall within-subject design on physiological measures, we extracted the regions with between-task significant differences of blinks and pupillary response for each participant, and also calculated the between-task differences for each participant in their subjective ratings, and then performed multiple regressions on these between-task differences. For blinks, we found that between-task blink differences are not predicted by any of the subjective ratings.

For pupillary response, in a best-fit multiple regression model with “Discontinuity”, “Happy”, “Bored” and “Excited” as predictors, we found that “Discontinuity” (b = 0.43, t(36) = -2.92, p < .01), “Excited” (b = -0.63, t(36) = -2.7, p < .05) and “Happy” (b = 0.40, t(36) = 1.79, p = .08) significantly predicted the differences between pupillary responses across the two kinds of tasks, with the model explained a significant proportion of variance in the differences (adjusted R^2 = 0.30, F(4, 36) = 5.26, p < .005). Variance inflation factors for each predictor in the model were examined, and no multicollinearity problems were found (square roots of VIF: “Discontinuity” at 1.11, “Excited” at 1.77, “Happy” at 1.70).

To summarize, the experience of solving puzzle problems as a discontinuous thinking process appeared to lead to larger pupillary dilation responses compared to routine problems. Solving puzzle problems as a happier experience and a possibly less “excited” experience also appeared to contribute to the higher pupillary activities during the puzzle problems.

4.0 GENERAL DISCUSSION

4.1 MAIN FINDINGS

Between-task differences in ratings were primarily found in problem solving related ratings, rather than in mood ratings. Ratings of moods in session one and session two reflect that "Frustrated" was the most significantly different between-tasks rating. Other ratings also showed slight between-task differences but the differences were not consistent across sessions. Ratings related to problem solving experience and solution perception were consistently different between tasks for both sessions, providing a strong manipulation check. Ratings in decreasing order of effect size are: "Discontinuity", "Cluelessness", "Suddenness" and "Obviousness" with ratings on puzzle problems higher in all these ratings.

Comparisons of physiological measures between the two kinds of tasks around solution time have confirmed many of our hypotheses. Around the time that participants press the button to indicate a solution was found, there were intermittently lower blink activities during puzzle problems for a 7s to 15s period. There were also larger pupillary dilations during puzzle problems for a 7s to 12s period around the time of solution declaration. For each kind of task, comparisons of z-scored blink frequency waveform and pupillary response waveform confirmed that during puzzle problems, there was a shorter period of relative higher pupillary activities prior to the time of solution declaration, and that the same period of higher relative pupillary activities was much longer prior to the solution declaration time during our routine problems.

Regression analyses results were consistent with our model that mood and problem solving ratings are mediators that predict solution time and pupillary responses between these two kinds of tasks. Ratings of “Obviousness” effectively predict solution time for puzzle problems. Differences in ratings “Discontinuity” and “Happy” effectively predict the differences that there were higher pupillary activities during puzzle problem tasks, while ratings of “Excited” predicts some aspects why some time pupillary activities could possibly be lower in puzzle problems.

4.2 INTERPRETATIONS AND IMPLICATIONS

4.2.1 Moods along the course of discoveries

In the literatures on insight problem solving, there has not been much discussion about the affective nature of insight during puzzle problem solving and discoveries (Gick & Lockhart, 1995). This is probably not due to a lack of interest about the topic, but because emotion is not a well investigated area in cognitive psychology before Ekman (1992). Also, within the field of higher-level cognition, there are a variable of perspectives about insight, ‘Aha!’ phenomena, and discoveries, let alone these researchers may not be familiar with the field of emotions at the same time. Among those that have looked into the relationships between affects and insight problem solving (Gick & Lockhart, 1995; Gruber, 1995), the feelings of suddenness and happiness have been discussed. Our major finding about mood ratings was that frustration may play an important role to distinguish discoveries from routine problems, and this is especially true for those difficult discoveries. Difficult puzzles or insight problems have a high chance of leading to moments of impasse because they likely prompt unworkable problem conceptualizations. In the picture, when a problem solver runs out of ideas about solution moves, frustration almost becomes inevitable. Gruber (Gruber, 1995) referred to various historical sources that describe the development of scientific ideas by some well-known scientists, such as Darwin and Einstein. He mentioned a few records, such as that found in the *Productive Thinking* (Wertheimer, 1959), which describes long episodes of doubt, despair, depression, and frustration during these courses of important scientific discoveries. In contrast to the alternative notion that insight is an intrinsically delightful experience, this characterization about the affective nature of insight points to the major reason for delight during an insightful moment. That is, it posits that an insight functions in a way similar to a pain reliever, and offers a relief from the unpleasantness experienced during getting-stuck in a problem solving or a difficult thinking process.

4.2.2 Characteristics of insight problems

Weisberg (Weisberg, 1995) suggested a dichotomic taxonomy to insight problems according to the definition proposed by Gestalt psychologists. The Gestalt psychologists (e.g., Kohler and Wertheimer) proposed that under certain circumstances, organisms could achieve solution of a problem through analysis of the problem even though there had not been extensive experience in the problem situation. Their claims were in response to Thorndike's (Thorndike, 1911) conclusion that all behaviour was based on trial and error. Weisberg proposed this taxonomy based on problem classification despite other researchers (Davidson, 1995; Gick & Lockhart, 1995) positing that the experience of suddenness and surprise should be the hallmarks of an insight because his examinations on the problems utilized heretofore in the investigations of insights are not solved using insights in a classical sense. In his taxonomy, the most important criteria are whether a problem solving experience involves discontinuity, and then followed by whether a problem involves restructuring.

Rather than having ourselves judge and claim the problems to be insight problems prior to the experiments, we incorporated these criteria into subjective ratings: the question in ratings of “Discontinuity” is about whether discontinuity is subjectively involved during the course of solution process, the question in ratings of “Cluelessness” is about whether there was any change of problem representation and perspective, in other words is restructuring involved. Apart from having these two criteria implemented as subjective ratings, we also looked into the suddenness and surprise nature of insight through subjective ratings: “Suddenness” ratings represented whether a problem solution came suddenly, and “Obviousness” ratings represented whether a problem solution is out of expectation or not, that is as a surprise or not. Our findings in this set of ratings about the subjective perspective of the problem solving process and solution perception revealed that our puzzle problems are significantly rated more insightful than the routine problems in all of these aspects. Among these rating differences, our results confirmed Weisberg claims that “Discontinuity” followed by “Cluelessness” indeed were the most effective differences from a problem solver first person perspective. We have also confirmed from a more mainstream perspective that our puzzle problems are more insightful in the sense that their solution perceptions were rated higher in their “Suddenness” and lower in their “Obviousness”.

Also, results in problem related ratings suggested that in session two, instead of just in session one, our participants also found the puzzle problems discontinuous and needed some changes in perspectives, and the solutions were less obvious and they came suddenly. This suggested that many of them experienced moments of insight in session two. We realized performance gains in session two as proposed by Knoblich et. al. (Knoblich et al., 1999; Knoblich et al., 2001), but there were also moments of insights due to rediscovering of solution mechanisms for puzzle problems instead of making totally new discovery as in session one.

One caveat about our design is that we used retrospective ratings as our subjective ratings. These rating questions were answered more than 15s after the button press of the time of solution declaration. Therefore, memory about the past problem solving experience when answering these retrospective questions might be less accurate than a metacognitive response right during the time of problem solving. We tried to avoid any other cognitive demand to disturb the insight problem solving process, and as a result, we could not utilize immediate metacognitive reports. A possible direction to solve this dilemma is to look at facial expression protocols of the problem solver during the tasks. However, at the present time, none of our mood ratings maps to known distinctive facial expression pattern because none of them belongs to the group of basic emotions. Mapping our problem solving experience and solution perception ratings to distinctive facial expression may prove to be even more challenging.

4.2.3 Blinks and pupillary profiles that reflects the temporal cognitive demands of insight problems

It is suggested that blink bursts reflect releases of resources used in stimulus-related cognition (Ohira et al., 1998), following high cognitive load (Fukuda, 1994; Ohira, 1996), or information processing (Ichikawa & Ohira, 2004). Effective preparation for externally-cued information has been proposed to suppress eyeblinks (Ohira, 1995). Our comparisons between blink frequency waveforms of puzzle problems with that of routine problems confirmed our hypotheses about cognitive load that there is a higher on-going load when the participants are solving the puzzle problems, leading to a lower blink frequency for the tasks. The period with lower blink frequency around button press was longer in session one than that in session two possibly because participants are more used to the puzzle problems in session two than in session one because all the solution mechanisms are completely new in session one. This leads to higher demands of cognitive resources in session one than in session two.

Pupillary changes can provide a reliable source of information about the cognitive demand and time course of information processing (Kahneman, 1973). Pupillary dilations were observed during conditions of heighten memory use, attentional allocation, or interpretation of more difficult material, and the dilations would persist if there is a sustained demand (Beatty, 1982). For instance, studies have shown that pupil dilation increases proportionally to the numbers of digits a person has to memorize (Granholm et al., 1996; Kahneman & Beatty, 1966). Our hypothesis that pupillary dilations are larger at the time of solution was supported, suggesting that solutions for the puzzle problems require higher cognitive demand and at the same time leading to a possibly higher arousal level. This period of higher cognitive demand followed till 2s to 7s after solution declarations, which also confirmed our prediction that there is a higher demand follow-through period for processing these solutions.

However, there is a caveat that these differences in pupillary and blink profiles are linked to highly comparable tasks with the puzzle problem being a single step puzzle problem. Studies on multiple-step problems are much more difficult to design if eye data need to be in synchrony to the external events because of larger variation in solution time and a much larger problem space for exploration. Detailed investigations in the possibilities need to be made.

4.2.4 An emerging sense of clarity as insight experience

The comparison of z-scored blinks and pupillary waveforms bring us a new interpretation of the insight experience. With externally driven response tasks, Siegle et al. (2008) demonstrated that a higher level of blink activities precede the processing of stimuli, followed by higher level of pupillary activities during which stimuli are processed, and then followed by another higher level of blink activities that comes after the stimulus processing. Our results suggested that for the puzzle problems, there was a cognitive processing event happening at about 6s prior to the time of solution declaration. Unlike in their study, this cognitive processing event at 6s before solution declaration is an internally driven mental event. We posit that this mental event is the time when an insight starts to emerge.

One may wonder why there is a 6s lag before solution declaration. First, we asked our participants try not to make mistake to declare a wrong solution correct, and this possibly required some time for them to double-check their proposed solution. Second, as suggested by our finding, we propose that the actual occurrence of an insight is very possibly a short moment with a gradual emerging sense of clarity about a solution instead of an all-sharp abrupt mental event.

We do agree that there is a sudden element in insight that the change of representation and perspective can occur in a short time. However, what we want to point out is that it takes time to go from a changed perspective to recognition of elements of solution in the new representation, and then the realization of the solution. The moments of solution declaration from a moment of insight may be very dramatic, such as when Archimedes apparently ran naked through the street to celebrate his discovery. We proposed that this stage of an emerging sense of clarity about a solution takes time. The more complex the problem, the longer it is going to take to gain complete clarity. For example, it still takes a second or so for us to bring all elements of a Necker cube or the Maiden and Witch figure to a coherent perspective to let us has a sense of clarity that we have recognized an alternative figure in the picture. However, along the whole process of insightful discovery, our memory and impression usually dwell in the moments of clarity and solution declaration.

We are not the first one to have this emerging characterization of insight. Gruber (1995) pointed out that historical scientific discoveries comprised of stages of developments that involved variations in intuitions and affective climates. He was opposed to the claim made by the majority of researchers on insight that a large amount of important scientific discoveries are made in some all-of-a-sudden manner as if these moments took no time. He pointed out that people made a number of mistakes in the characterization of historical materials about these discoveries. There are commonly losses of memory regarding intermediate details, post hoc rationalizations for some empirical accounts, and decontextualizations when people process and think about these hallmark discoveries. More accurate and reliable historical records always reveal that there were stages in the development of these discoveries, and there were never a real sharp leap into something totally new.

In our current puzzle problem, there are a few steps that a person needs to take to proceed to the solution declaration stage. The participants would need to recognize the correct move, and then re-evaluate if the final expression after the move will be a valid and correct expression. We should be cautious here that the higher level pupillary activities before solution declaration can possibly be shorter than 6s if a puzzle problem requires a shorter solution recognition and confirmation time. To confirm this, studies need to be conducted using another single-step insight puzzle problem that has a shorter or longer solution recognition and confirmation time.

Schooler et al. (Schooler, Fallshore, & Fiore, 1995) used vision and navigation to compare to the two perspectives of insight problem solving: the restructuring theories and the heuristic search theories. They posited that it is possible to see memories and ideas that are important to the solution as objects in the problem space and problem solving is basically navigation in the space to get them. It takes time to search for the right space, and it takes time to recognize a solution.

4.3 COMPARISON TO THE DEPICTION OF INSIGHT BY METCALFE & WIEBE

With these limitations of our study in mind, there are still a few major advances in the perspective of insight problem solving from the classic point of view by Metcalfe and Wiebe (1987). First, from regression results, we found that puzzle problems and routine problems led to different pupillary dilations with our perception and mood about the tasks as possible mediators of the effect. Profiles of pupillary responses and blinks are found to be indicators that depict characteristics of the task nature when a person is unable to predict his/her own problem solving performance on an insight problem. Second, we confirmed that conventional suddenness and surprise elements of insight are important characteristics of insight problem solving events; however, we found that another two characteristics that have been mentioned by a number of researchers (e.g. Knoblich et al. 1999; Weisberg, 1995) and the older Gestalt psychologists – discontinuity and restructuring are possibly more important in characterization an insight problem solving event. Third, instead of the traditional sharp and abrupt but indistinct characterization of the insight moment, our data suggested that an insight is possibly an emerging moment of a sense of clarity about the solution which takes time before the time of an ‘Aha!’ solution declaration, which converges with some of the claims made according to records of historical scientific discoveries (Gruber, 1995). Further work about mapping moments of insight on other classic insight problems can be carried out using the same technique of eye tracking. It is just the beginning of an era that we are able to characterize the true nature of insight moments to greater details.

APPENDIX A. MATCH-STICK PUZZLE PROBLEMS AND ANSWERS

A.1 SESSION ONE

Problem	Solution
X – VIII = IV	X**I** – VII = IV
III + VII = XII	III + VII**I** = XI
XIII = VII – V	XII = VII **+** V
IX = XIII – VI	IX = XIII - **I**V
X – XIII = II	X – **V**III = II
VIII = VII – VII	VII = VII **=** VII

VII = I – VI	VII – I **=** VI
IX + III = V	IX – III = V**I**

A.2 SESSION TWO

Problem	**Solution**
IX – IV = VII	X**I** – IV = VII
VI – VIII = II	VI **=** VIII – II
X = IX – II	X = IX **+** I
X = IX – II	X = IX **+** I
X = XIII + IV	**I**X = XIII – IV

XI + VII = XIII	**V**I + VII = XIII
IV + VIII = X	IV + VII = X**I**
XII – VII = II	XI – VII**I** = III

APPENDIX B. ROMAN NUMERAL ROUTINE PROBLEMS AND ANSWERS

B.1 SESSION ONE

Problem	Expression With Missing Numeral	Number to Report
XII + XVI = ?	XII + XVI = XXVIII	13
? = XXXII – XIX	XIII = XXXII – XIX	8
XXVII – ? = XI	XXVII – XVI = XI	8
XXIV = ? + XIX	XXIV = V + XIX	6
? + XVII = XXVI	IX + XVII = XXVI	7
VIII = XXIII – ?	VIII = XXIII – XV	7
? – XXII = XIII	XXXV – XXII = XIII	11

XXXIII = ? + XVII	XXXIII = XVI + XVII	9

B.2 SESSION TWO

Problem	Expression With Missing Numeral	Number to Report
XXVIII – XIX = ?	XXVIII – XIX = IX	6
? = XXIV + VII	XXXI = XXIV + VII	11
VII + ? = XVIII	VII + XI = XVIII	7
IX = ? – XII	IX = XXI – XII	8
? – XVII = XIV	XXXI – XVII = XIV	10
XXXII = XIX + ?	XXXII = XIX + XIII	9
? + XVII = XXXV	XVIII + XVII = XXXV	11
XV = ? – XVI	XV = XXXI – XVI	10

APPENDIX C. SUBJECTIVE RATING QUESTIONS

C.1 MOOD

C.1.1 "Frustrated"

I was feeling VERY NOT FRUSTRATED in working on this problem	I was feeling VERY FRUSTRATED in working on this problem

1 **2** **3** **4** **5** **6** **7**

C.1.2 “Happy”

I was feeling VERY NOT HAPPY	I was feeling VERY HAPPY
in working on this problem	in working on this problem

1 2 3 4 5 6 7

C.1.3 “Excited”

I was feeling VERY NOT EXCITED	I was feeling VERY EXCITED
in working on this problem	in working on this problem

1 2 3 4 5 6 7

C.1.4 “Bored”

I was feeling VERY NOT BORED	I was feeling VERY BORED
in working on this problem	in working on this problem

1 2 3 4 5 6 7

C.2 EXPERIENCE OF PROBLEM SOLVING PROCESS

C.2.1 "Cluelessness"

Early on in this problem...... I had a general plan for how to work on the solution	Early on in this problem...... I had no clue as to how to solve the problem

1 2 3 4 5

C.2.2 "Discontinuity"

Early on in this problem...... I thought about the problem in basically the same way as at the end	Early on in this problem...... I thought about the problem in a very different way than in the end

1 2 3 4 5

C.3 SOLUTION PERCEPTION

C.3.1 "Obviousness"

When I thought of the solution...... I had to try out to see if it is correct				When I thought of the solution...... I instantly know it is correct
1	**2**	**3**	**4**	**5**

C.3.2 "Suddenness"

When I thought of the solution...... I had slowly built it up in pieces				When I thought of the solution...... it had come to me all of a sudden
1	**2**	**3**	**4**	**5**

BIBLIOGRAPHY

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: an activation-based model. *Cognitive Science: A Multidisciplinary Journal, 26*(1), 39 - 83.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, N.J.: L. Erlbaum Associates.

Anderson, J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the Cost of Goal Retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1331-1346.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30*, 221-256.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). San Diego, CA: Academic Press.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*, 723-767.

Beatty, J. (1982). Task-evoked pupillary responses processing load and the structure of processing resources. *Psychological Bulletin, 91*, 276–292.

Broadbent, D. E. (1971). *Decision and stress*. London, New York,: Academic Press.

Christensen, B. T., & Schunn, C. D. (2005). Spontaneous Access and Analogical Incubation Effects. *Creativity Research Journal*, *17*(2), 207-220.

Christensen, B. T., & Schunn, C. D. (2007). The relationship of analogical distance to analogical function and preinventive structure: The case of engineering design. *Memory & Cognition*, *35*(1), 29-38.

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(1), 14-27.

Darwin, C. (1872). *The expression of emotion in man and animals*. London: Murray.

Davidson, J. E. (1995). The suddenness of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight* (pp. 125-155). Cambridge: The MIT Press.

Dominowski, R. L., & Dallob, P. (1995). Insight and problem solving. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge: The MIT Press.

Duncker, K. (1945). On problem solving. *Psychologicla Monographs*, *58*(5, Whole No. 270).

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3), 169-200.

Ericsson, K. A., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review*, *102*(2), 211-245.

Eysenck, M. W. (1982). *Attention and arousal, cognition and performance*. Berlin ; New York: Springer-Verlag.

Falkenhainer, B., Forbus, K., & Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence, 41*, 1-62.

Fogarty, C., & Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. *International Journal of Psychophysiology, 8*(1), 35-42.

Fukuda, K. (1994). Analysis of eyeblink activity during discriminative tasks. *Perceptual & Motor Skills, 79*(3 Pt 2), 1599-1608.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

Gentner, D., Forbus, K. D., & Law, K. (1995). MAC/FAC: A Model of Similarity-based Retrieval. *Cognitive Science, 19*(2), 141-205.

Gentner, D., Holyoak, K. J., & Kokinov, B. K. (Eds.). (1998). *The Analogical Mind*. Cambridge: MIT Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1-38.

Gick, M. L., & Lockhart, R. S. (1995). Cognitive and affective components of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge: The MIT Press.

Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology, 33*(4), 457-461.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*, 462-466.

Gruber, H. E. (1995). Insight and affect in the history of science. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge: The MIT Press.

Holyoak, K., & Thagard, P. (1997). The Analogical Mind. *American Psychologist, 52*(1), 35-44.

Ichikawa, N., & Ohira, H. (2004). Eyeblink activity as an index of cognitive processing: temporal distribution of eyeblinks as an indicator of expectancy in semantic priming. *Perceptual & Motor Skills, 98*(1), 131-140.

Janisse, M. P. (1976). The relationship between pupil size and anxiety: A review. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 3). Washington, D. C.: Hemisphere.

Janisse, M. P. (1977). *Pupillometry : the psychology of the pupillary response*. New York: Hemisphere Pub. Corp.

Kahneman, D. (1971). Changes in pupil size and visual discriminations during mental activity. In J. R. Pierce & J. R. Levine (Eds.), *Visual science*. Bloomington: Indiana University Press.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(1583-1585).

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*(3), 374-419.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1534–1556.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*(7), 1000-1009.

Köhler, W. (1927). *The mentality of apes* (2d. ed.). London: Routledge & Kegan Paul.

Köhler, W. (1921). *Intelligenzprüfungen an Menschenaffen* (2nd durchgesehene Aufl der "Intelligenzprüfungen an Anthropoiden I" ed.). Berlin: J. Springer.

Loewenfeld, I. E., & Lowenstein, O. (1999). *The pupil : anatomy, physiology, and clinical applications*. Boston: Butterworth-Heinemann.

Lovett, M. C., & Schunn, C. D. (1999). Task Representations, Strategy Variability, and Base-Rate Neglect. *Journal of Experimental Psychology: General, 128*(2), 107-130.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 176-201.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15*(3), 238-246.

Namy, L. L., & Gentner, D. (2002). Making a Silk Purse Out of Two Sow's Ears: Young Children's Use of Comparison in Category Learning. *Journal of Experimental Psychology: General, 131*(1), 5-15.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, N. J.: Prentice Hall.

Ohira, H. (1995). Analysis of eyeblink activity during self-referent information processing in mild depression. *Perceptual & Motor Skills, 81*(2), 1219-1229.

Ohira, H. (1996). Eyeblink activity in a word-naming task as a function of semantic priming and cognitive load. *Perceptual & Motor Skills, 82*(1), 835-842.

Ohira, H., Winton, W. M., & Oyama, M. (1998). Effects of stimulus valence on recognition memory and endogenous eyeblinks: Further evidence for positive-negative asymmetry. *Personality and social psychology bulletin, 24*, 986-993.

Ohlsson, S. (1984). Restructuring revisited: I. Summary and critique of the Gestalt theory of problem solving. *Scandinavian Journal of Psychology, 25*, 65-78.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1-44). New York: Harvester Wheatsheaf.

Schooler, J. W., Fallshore, M., & Fiore, S. M. (1995). Epilogue: Putting insight into perspective. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge: The MIT Press.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*(2), 166-183.

Siegle, G. J., Granholm, E., Ingram, R., & Matt, G. (2001). Pupillary response and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry, 49*, 624-636.

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45*(5), 679-687.

Simon, H. A., & Chase, W. G. (1973). Skill in chess. *American Scientist, 61*(4), 394-403.

Stern, J. A., Walrath, L. C., & Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology, 21*(1), 22-33.

Thorndike, E. L. (1911). *Animal intelligence*. New York: Macmillan.

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge: The MIT Press.

Wertheimer, M. (1925). *Drei Abhandlungen zur Gestalttheorie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Wertheimer, M. (1959). *Productive thinking* (Enl. ed.). New York: Harper.